THE UNIVERSITY *of* EDINBURGH

Phenome-wide association study (PheWAS) on the genetic determinants of serum urate level and disease outcomes in UK Biobank

Xue Li

THE UNIVERSITY
of EDINBURGH

Doctor of Philosophy – The University of Edinburgh – 2018

DECLARATION

I, Xue Li hereby declare that the research within this thesis is my own work. I conducted all aspects of this research and any assistance received has been duly acknowledged. The work described has not been submitted for any other degree or professional qualification.

Signature:…… ….. Date:.…….……..…..

AKNOWLEDGMENTS

This thesis would not have been completed without the help and guidance from a group of people. They all have been encouraging and supportive during my study in the University of Edinburgh. I would like to express my sincere thankfulness to all of them.

I wish to give my deepest thankfulness to my supervisors **Dr. Evropi Theodoratou** and **Professor Harry Campbell** for their patient guidance. They gave me tremendous support and encouragement throughout my PhD and provided me opportunities to be well trained and learn new knowledge. They can always manage to match their breadth of scientific knowledge with kind advice to inspire me to strive toward excellence. Their impressive kindness and understanding provide me a pleasure study experience in the past three years.

I would like to specially thank **Dr. Athina Spiliopoulou**, **Dr. Maria Timofeeva**, and **Dr. Peter Joshi** for their guidance and advice on the statistical issues regarding this thesis. I wish to thank **Professor Paul Mckeigue**, **Professor Jim Wilson**, **Professor Alan C. Wright**, and **Professor Igor Rudan** for their support. I wish to thank **Professor John P.A. Ioannidis** from Stanford University for his guidance. I would like to acknowledge **Professor Joshua C. Denny**, **Dr. Wei-qi Wei** and **Dr. Aliya Gifford** from the Vanderbilt University Medical Centre for their help with the data analysis. I wish to thank my teammate **Xiangrui Meng** for his support and help.

I would also like to acknowledge all members in the **Usher Institute of Population Health Sciences and Informatics** and the **Institute of Genetics and Molecular Medicine (IGMM)** for their academic and technical support.

I wish to acknowledge the **China Scholarship Council** for funding me and giving me the opportunity to study in the UK.

I would like to thank my friends **Peige Song**, **Xin Wang**, **Tian Yang**, **Xiangrui Meng**, **Yazhou He**, **Hongjiang Wu** and **You Li** in the UK and all my friends in China for their care and company.

Finally, I wish to thank my beloved husband, my parents, my younger brother and sister for their enduring love, support and care.

ABSTRACT

Introduction

Elevated serum uric acid (SUA) concentration, known as hyperuricaemia, is a common abnormity in individuals with metabolic disorders. There is increasing evidence supporting the link between high SUA level and the increased risk of a wide range of clinical disorders, including hypertension, cardiovascular diseases (CVD), chronic renal diseases and metabolic syndrome. Although there are considerable research efforts in understanding the pathogenic pathways of high SUA level and the related clinical consequences, their causal relationships have not been established except for gout. Like other complex traits, genetic determinants play a substantial role (an estimated heritability of 40-70%) in the regulation of SUA level. Investigating the role of genetic variants related to SUA in various diseases might provide evidence for the above hypothesis which links uric acid to clinical disorders.

Method

Umbrella review was carried out first to provide a comprehensive overview on the range of health outcomes in relation to SUA level by incorporating evidence from systematic reviews and meta-analyses of observational studies, meta-analyses of randomised controlled trials (RCTs), and Mendelian randomisation (MR) studies. The umbrella review summarised the range of related health outcomes, the magnitude, direction and significance of identified associations and effects, and classified the evidence into four categories (class I [convincing], II [highly suggestive], III [suggestive], and IV [weak]) with assessment of multiple sources of biases.

Then, a MR-PheWAS (Phenome-wide association study incorporated with Mendelian randomisation [MR]) was performed to investigate the associations between the 31 SUA genetic risk variants and a very wide range of disease outcomes by using the interim release data of UK Biobank (n=120,091). The SUA genetic risk loci were employed as instruments individually. The framework of phenome was defined by the PheCODE schema using the International Classification of Diseases (ICD) diagnosis codes documented in the health records of UK Biobank. Phenome-wide association test was performed first to identify any association across the SUA genetic risk loci and the phenome; MR design and HEIDI (heterogeneity in dependent instruments) tests were then applied to distinguish the PheWAS associations that were due to causality, pleiotropy or genetic linkage.

To validate the MR-PheWAS findings, an enlarged Phenome-wide Mendelian randomisation (PWMR) analysis were performed by using data from the full UK Biobank cohort (n=339,256). A weighted polygenic risk score (GRS), incorporating effect estimates of multiple genetic risk loci, was employed as a proxy of the SUA level. The framework of phenome was defined by both the PheCODE schema and an alternative Tree-structured phenotypic model (TreeWAS) for analysis. Significant associations from these analyses were taken forward for replication in different populations by analysing data from various GWAS consortia documented in the MR-base database. Sensitivity analyses examining the pleiotropic effects of urate genetic risk loci on a set of metabolic traits were performed to explore any causal effects and pleiotropic associations.

Results

The umbrella review included 101 articles and comprised 144 meta-analyses of observational studies, 31 meta-analyses of randomised controlled trials and 107 Mendelian randomisation studies. This remarkable assembly of evidence explored 136 unique health outcomes and reported convincing (class I) evidence for the causal role of SUA in gout and nephrolithiasis. Furthermore, highly suggestive (class II) evidence was reported for five health outcomes, in which high SUA level was associated with increased risk of heart failure, hypertension, impaired fasting glucose or diabetes, chronic kidney disease, and coronary heart disease mortality in the general population. The remaining 129 associations were classified as either suggestive or weak.

The MR-PheWAS (using the interim release cohort) identified 25 disease groups/ outcomes to be associated with SUA genetic risk loci after multiple testing correction ($p<8.6 \times 10^{-5}$). The MR IVW (inverse variance weighted) analysis implicated a causal role of SUA level in three disease groups: inflammatory polyarthropathies (OR=1.22, 95% CI: 1.11 to 1.34), hypertensive disease (OR=1.08, 95% CI: 1.03 to 1.14) and disorders of metabolism (OR=1.07, 95% CI: 1.01 to 1.14); and four disease outcomes: gout (OR=4.88, 95% CI: 3.91 to 6.09), essential hypertension (OR=1.08, 95% CI: 1.03 to 1.14), myocardial infarction (OR=1.16, 95% CI: 1.03 to 1.30) and coeliac disease (OR=1.41, 95% CI: 1.05 to 1.89). After balancing pleiotropic effects in MR Egger analysis, only gout and its encompassing disease group of inflammatory polyarthropathies were considered to be causally associated with SUA level. The analysis also highlighted a locus (*ATXN2/S2HB3*) that may influence SUA level and multiple cardiovascular and autoimmune diseases via pleiotropy.

The PWMR analysis, using data from the full UK Biobank cohort (n=339,256), examining the association with 1,431 disease outcomes, identified 13 phecodes that were associated with the weighted GRS of SUA level with the p value passing the significance threshold of PheWAS ($p<3.4\times10^{-4}$). These phecodes represent 4 disease groups: inflammatory polyarthropathies (OR=1.28; 95% CI: 1.21 to 1.35; $p=4.97\times10^{-19}$), hypertensive disease (OR=1.08; 95% CI: 1.05 to 1.11; $p=6.02\times10^{-7}$), circulatory disease (OR=1.05; 95% CI: 1.02 to 1.07; $p=3.29\times10^{-4}$) and metabolic disorders (OR=1.07; 95% CI: 1.03 to 1.11; $p= 3.33\times10^{-4}$), and 9 disease outcomes: gout (OR=5.37; 95% CI: 4.67 to 6.18; $p= 4.27\times10^{-123}$), gouty arthropathy (OR=5.11; 95% CI: 2.45 to 10.66; $p=1.39\times10^{-5}$), pyogenic arthritis (OR=2.10; 95% CI: 1.41 to 3.14; $p=2.87\times10^{-4}$), essential hypertension (OR=1.08; 95% CI: 1.05 to 1.11; $p=6.62\times10^{-7}$), coronary atherosclerosis (OR=1.10; 95% CI: 1.05 to 1.15; $p=1.17\times10^{-5}$), ischaemic heart disease (OR=1.10, 95% CI: 1.05 to 1.15; $p=1.73\times10^{-5}$), chronic ischaemic heart disease (OR=1.10, 95% CI: 1.05 to 1.15; $p=1.52\times10^{-5}$), myocardial infarction (OR=1.15, 95% CI=1.07 to 1.23, $p=5.23\times10^{-5}$), and hypercholesterolaemia (OR=1.08, 95% CI: 1.04 to 1.13, $p=3.34\times10^{-4}$). Findings from the TreeWAS analysis were generally consistent with that of PheWAS, with a number of more sub-phenotypes being identified. Results from IVW MR suggested that genetically determined high serum urate level was associated with increased risk of gout (OR=4.53, 95%CI: 3.64-5.64, $p=9.66\times10^{-42}$), CHD (OR=1.10, 95%CI: 1.02 to 1.19, p=0.009), myocardial infarction (OR=1.11, 95%CI:1.02 to 1.20, p=0.011) and decreased level of HDL-c (OR=0.93, 95%CI:0.88 to 0.98, p=0.004), but had no effect on RA (OR=0.92, 95%CI: 0.84 to 1.01, p=0.085) and ischaemic stroke (OR=1.03, 95%CI: 0.93 to 1.14, P= 0.582). Egger MR indicated pleiotropic effects on the causal estimates of DBP ($P_{pleiotropy} = 0.014$), SBP ($P_{pleiotropy} = 0.003$), CHD ($P_{pleiotropy} = 0.008$), myocardial infarction ($P_{pleiotropy} = 0.008$) and HDL-c ($P_{pleiotropy} = 0.016$). When balancing out the potential pleiotropic effects in Egger MR, a causal effect can only be verified for gout (OR=4.17, 95%CI: 3.03 to 5.74, $P_{effect} = 1.27 \times 10^{-9}$; $P_{pleiotropy} = 0.485$). Sensitivity analyses on the GRSs of different groups of pleiotropic loci support an inference that pleiotropic effects of genetic variants on urate and metabolic traits contribute to the observed associations with cardiovascular/metabolic diseases.

Conclusions

This thesis presents a comprehensive investigation on the health outcomes in relation to SUA level. The causal relationship between high SUA level and gout is robustly verified in this thesis with consistent evidence from the umbrella review, the MR-PheWAS and the PWMR. The association of high SUA level with hypertension and heart diseases is supported by both

the evidence from umbrella review and analyses conducted in this thesis, however, given the caveat of pleiotropy in the causal inference, a conclusion of causality on hypertension and heart diseases is not robust enough based on the current findings. Furthermore, the epidemiological evidence from the umbrella review indicated that high SUA level was associated with several components of metabolic disorders, and the analyses of the UK Biobank data identified a significant association with metabolic disorders and a sub-phenotype (hypercholesterolaemia). The causal inference in this study is limited by the common difficulty of pleiotropy caused by the use of multiple genetic instruments. Although we have performed sensitivity analysis by excluding the key pleiotropic locus, unmeasured pleiotropy and biases are still possible. In particular, unbalanced pleiotropy is recognised as an issue for the causal connections on the association between SUA level and hypertension. Other potential causal relevance of SUA level with respiratory diseases and ocular diseases is also worthy of further investigation. Overall, when taken together the findings from umbrella review, MR-PheWAS, PheWAS/TreeWAS analysis, MR replication and sensitivity analysis conducted in this thesis, I conclude that there are robust associations between urate and several disease groups, including gout, hypertensive diseases, heart diseases and metabolic disorders, but the causal role of urate only exists in gout. This study indicates that the observed associations between urate and cardiovascular/metabolic diseases are probably derived from the pleiotropic effects of genetic variants on urate and metabolic traits. Further investigation of therapies targeting the shared biological pathways between urate and metabolic traits may be beneficial for the treatment of gout and the primary prevention of cardiovascular/metabolic diseases.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

LIST OF FIGURES

Chapter 1

Chapter 3

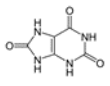
Chapter 4

Chapter 5

1 BACKGROUND

1.1 Uric acid

1.1.1 Physiochemical properties

Uric acid ( , 7, 9-dihydro-1*H*-purine-2, 6, 8(3*H*)-trione) is a heterocyclic organic compound of carbon, hydrogen, nitrogen and oxygen ($C_5H_4N_4O_3$) with a molecular weight of 168 g/mol (1). Uric acid behaves as a weak hydrogenated acid with a dual dissociation constant (pK_α=5.75, pK_β=10.30) (1). The hydrolysis reaction (uric acid <=> $urate^- + H^+$) is prone to shift to the right in weakly alkaline environment. The water solubility of uric acid is relatively low (0.6 mg/100 mL, at 20 °C) varying according to environmental temperature and the pH (potential of hydrogen) (1).

In most mammals, uric acid as a break-down product of purine metabolism could be further metabolised into a more soluble allantoin that can be completely excreted via urine, therefore, the vast majority of mammalian species have a very low concentration of uric acid (range: 30-89 μmol/L) in the blood (2, 3). However, in humans and other hominoids, due to the evolutionary functional loss of uricase (an enzyme, catalysing the conversion of uric acid into allantoin), uric acid is unable to be further degraded and thereby exists as the final product of purine catabolism (3). Under normal physiological condition (i.e., pH 7.4 and 37°C), uric acid predominantly circulates as the urate anion combined with a variety of cations (e.g., sodium, potassium, calcium, ammonium, and magnesium) in the plasma, extracellular and synovial fluid, and only a very small proportion (less than 5%) is bound to serum albumin (4). Due to the high concentration of sodium in human body fluids, monosodium urate (MSU) monohydrate ($NaC_5H_3N_4O_3 \cdot H_2O$) exists as one of the most common forms of ionised urate, in which a urate molecule is bonded to one sodium and one water molecule (5). The solubility limit of urate in human blood is approximately 405 μmol/L (6.8 mg/dL) (5). The normal reference interval of serum urate level is 89-357 μmol/L (1.5-6.0 mg/dL) in women and 149-416 μmol/L (2.5-7.0 mg/dL) in men (6). When the serum urate level exceeds the solubility limit (at a status of near saturation or supersaturation), urate crystals will deposit as MSU preferentially in and around peripheral joints (5).

1.1.2 Metabolic homeostasis

The metabolic homeostasis of uric acid is determined by the balance between its production and elimination. This involves a variety of complicated biological processes, including hepatic purine catabolism, renal excretion and intestinal uricolysis. Under normal physiological circumstances, healthy adults would maintain a total uric acid body pool of approximately 1000mg (1200mg in males and 600-700mg in females), and achieve a daily turnover rate at 60% from the balance of uric acid metabolism (7).

1.1.2.1 Uric acid production

Uric acid is typically not ingested from the diet but produced from the degradation of exogenous and endogenous purines (human diet contains little urate but many purine precursors). The major site of uric acid production is the liver, with small amounts being produced in other tissues like intestine, muscle, kidney and vascular endothelium (8). Its production depends on the process of purine biosynthesis and degradation. Normally, the enzymes involved in the purine metabolism maintain a balanced ratio between purine synthesis and degradation in the cell.

- ***Purine de novo synthesis and salvage***

Purines, as the monomeric precursors of nucleic acids, perform many important biological functions in the cell. They function as essential components of DNA and RNA nucleic acids to store, transcribe and translate genetic information (9). Purines are also the structural components of some co-enzymes (e.g., nicotinamide adenine dinucleotide [NAD]) and provide the source of cellular energy through adenosine triphosphate (ATP). These functions have been shown to play critical roles in modulating cellular energy metabolism and signal transduction (10). A balanced quantity of purines is required by cells for their growth, proliferation and survival. In order to replace the obligatory loss of purines during tissue nucleic acid turnover, purine nucleotides are mainly supplied through two biosynthesis pathways: *de novo* synthesis and salvage (**Figure 1-1**).

Purine nucleotides can be synthesised *de novo* with simple starting materials (i.e., glycine, glutamine, aspartate, formate, and CO_2) which involves a multistep biosynthesis (11). Purine biosynthesis is initiated on a backbone of ribose-5-phosphate to form a phosphorylated ring structure (5-phosphoribosyl-1 Pyrophosphate, PRPP), and this reaction is catalysed by the PRPP aminotransferase. Following a series of reactions utilising ATP, glutamine, glycine and aspartate, this pathway yields inosine monophosphate (IMP). IMP is then converted into either adenosine nucleotide (AMP) or guanosine nucleotides (GMP). Of note, the process of purine *de*

novo synthesis is highly energy consuming, which consumes 8 ATP equivalents for the synthesis of AMP and 9 ATP equivalents for the synthesis of GMP. The two rate-limiting enzymes of these reactions, PRPP synthase and glutamine-PRPP amidotransferase, are both subject to the feed-back control of various purine nucleotides. The second pathway for purine biosynthesis is known as purine salvage, which contributes to recover the purine bases (i.e., adenine, hypoxanthine, and guanine), either from the nucleic acid turnover or from the dietary nucleic acids, into the forms of adenine and guanine nucleotides. Salvage process involves re-synthesis of nucleotides from bases through two phosphoribosyltransferases: (i) adenine phosphoribosyltransferase (APRT) mediating the conversion from adenine to AMP; and (ii) hypoxanthine-guanine phosphoribosyltransferase (HGPRT) acting on either hypoxanthine to produce IMP or on guanine to produce GMP. These enzymes involved in salvage pathway are widely distributed among human tissues. The *de novo* synthesis and salvage pathways contribute interactively to maintain a constant and desired purine nucleotide pool for human body.

Exogenous purines from diet, which are absorbed mainly as nucleosides and free bases, can also partly enter the body pool of purines. An abundant supply of exogenous purine precursors could affect the pathways of purine metabolism (12). Its effects on the *de novo* purine synthesis include: (i) depressing the activity of PRPP aminotransferase that catalyses the initial reaction of purine *de novo* synthesis, given this enzyme can be inhibited by the feed-back of the high level of purine nucleotides (e.g., ATP, ADP); and (ii) reducing the supply of PRPP (S-phosphoribosyl-1-pyrophosphate), a substrate for *de novo* synthesis. These actions subsequently divert the PRPP away from the *de novo* synthesis process resulting in the *de novo* synthesis being minimised or switched off. Moreover, the purine salvage process is also subjected to feedback inhibition by purine levels. Once the cellular level of purine nucleotides reaches the required level, an increasing load of purines will inhibit the activities of enzymes (i.e., APRT and HGPRT) and consequently direct the purines towards the degradation process to produce uric acid.

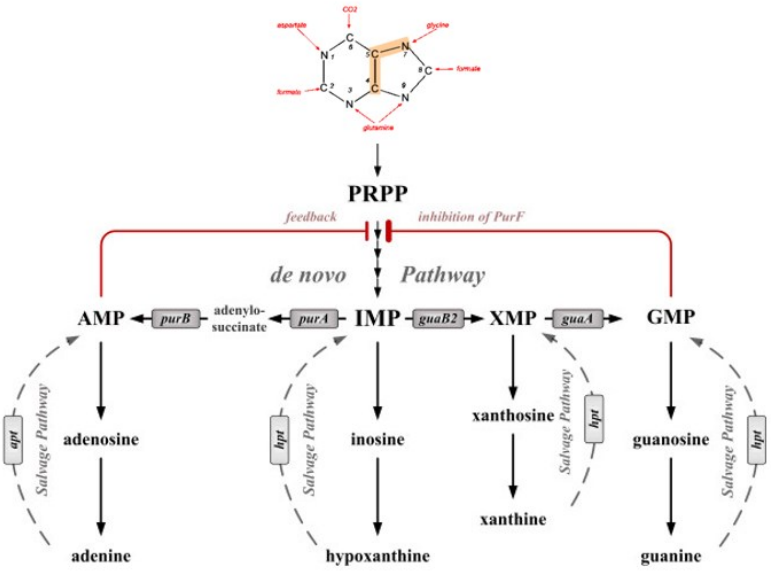


Figure 1 - 1: Precursors and pathway of purine *de novo* synthesis and salvage.

Dark grey boxes denote enzymes contributing to purine *de novo* synthesis pathway; light grey boxes denote enzymes contributing to purine salvage pathway. Red lines represent feedback inhibition of the rate limiting steps of the purine biosynthetic pathway, regulating the activity of the PRPP amidotransferase. Abbreviations: PRPP, 5-phosphoribosyl 1-pyrophosphate; IMP, inosine 5-monophosphate; AMP, adenosine 5-monophosphate; XMP xanthosine 5-monophosphate; GMP, guanosine 5-monophosphate; purA, adenylosuccinate synthetase; purB, adenylosuccinate lyase; guaB2, IMP dehydrogenase; guaA, GMP synthase; apt, adenine phosphoribosyltransferase; hpt, hypoxanthine-guanine phosphoribosyltransferase; (Source: adapted from reference (12) with permission covered by Creative Commons Attribution License [CC BY]).

- ***Purine degradation and uric acid formation***

The main pathways of purine degradation are outlined in **Figure 1-2**. Briefly, purine nucleotides are converted into nucleosides by intracellular nucleosidases first; nucleosides are then converted to inosines by adenosine deaminase; inosines are further degraded by the enzyme purine nucleoside phosphorylase (PNP) to release the purine base and ribose-l-P. The PNP products are merged into xanthine by guanine deaminase and xanthine oxidase, and xanthine is then oxidised to uric acid by this latter enzyme. The major steps of uric acid formation could be summarised as:

(i) Dephosphorylation – The purine molecules (AMP, IMP, and GMP) are dephosphorylated into the corresponding nucleotides (adenosine, inosine, and guanosine) by the enzyme 5-prime nucleotidase (5'-NT).

(ii) Deamination – The AMP (nucleotide) and adenosine (nucleoside) are deaminated into IMP and inosine correspondingly. These reactions are catalysed by AMP deaminase (AMPD) and adenosine deaminase (ADA), respectively.

(iii) Glycosidic bond cleavage – The nucleoside inosine is then converted into hypoxanthine and the nucleoside guanosine is converted into guanine, which are both catalysed by purine nucleotide phosphorylase (PNP).

(iv) Uric acid formation – Hypoxanthine and guanine are oxidised into xanthine by xanthine oxidase (XO) and guanine deaminase (GDA) respectively. Xanthine is finally catalysed into uric acid by xanthine oxidase.

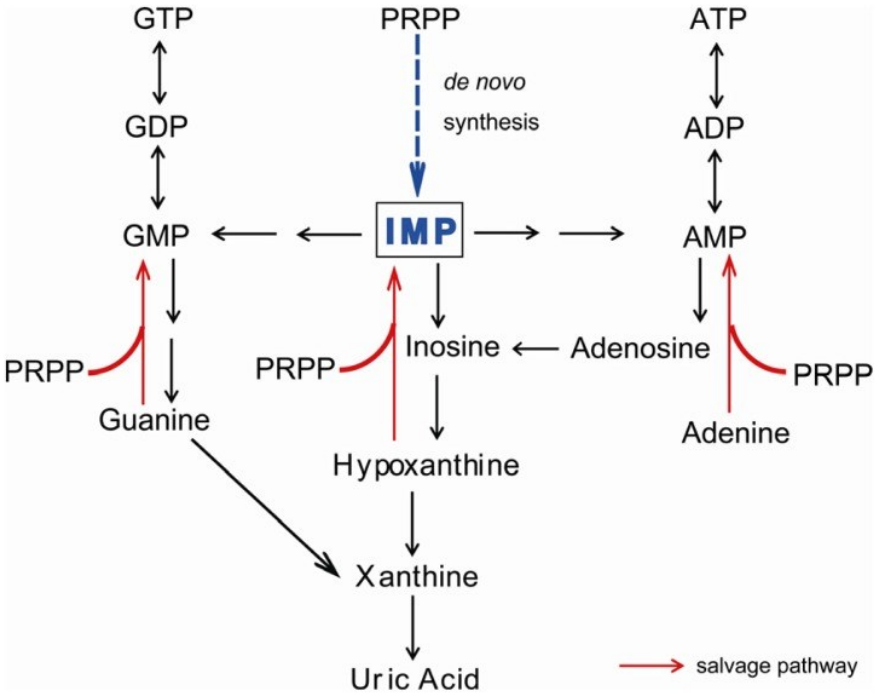


Figure 1 - 2: Purine degradation and UA production.

Both endogenously and exogenously purines share the same pathway for degradation to uric acid. Abbreviations: GTP, guanosine triphosphate; ATP, adenosine triphosphate; PRPP, 5-phosphoribosyl 1-pyrophosphate; IMP, inosine 5-monophosphate; AMP, adenosine 5-monophosphate; GMP, guanosine 5-monophosphate; (Source: adapted from Reference (13) with permission covered by CC BY).

- ***Inborn disorders in purine metabolism***

As explained above, each step of purine metabolism highly depends on the activities of enzymes catalysing the corresponding reactions. Genetic deficiency in any enzyme involved in purine synthesis, recycling or degradation processes, will cause purine nucleotides to not be

metabolised properly and result in different hereditary disorders (14). Inborn disorders resulting in abnormalities in purine *de novo* synthesis include: (i) phosphoribosylpyrophosphate synthetase superactivity, (ii) adenylosuccinase deficiency, and (iii) 5-amino-4-imidazolecarboxamide (AICA) riboside deficiency (AICA-ribosiduria). Inborn disorders causing abnormalities in purine salvage pathway include: (i) hypoxanthine-guanine phosphoribosyltransferase (HPRT) deficiency, and (ii) adenine phosphoribosyltransferase (APRT) deficiency. Inborn disorders leading to abnormalities in purine catabolism include: (i) muscle adenosine monophosphate (AMP) deaminase deficiency, (ii) adenosine deaminase deficiency, (iii) purine nucleoside phosphorylase deficiency, and (iv) xanthine oxidoreductase deficiency. These inborn disorders in purine metabolism would over-ride the control of uric acid metabolism homeostasis and result in either hypouricaemia or hyperuricaemia (see *Chapter 1, Section 1.2 "Hyperuricaemia"* for more information).

1.1.2.2 Uric acid excretion

The daily production of uric acid from purine catabolism is relatively constant at 300 to 400mg (8). Unlike other mammals having the enzyme uricase to convert uric acid into a more soluble allantoin for excretion, metabolism of uric acid in human tissues is negligible. To maintain homeostasis, uric acid is eliminated intact from the human body via two routes: the gastrointestinal tract and the kidney (15).

- ***Intestinal secretion***

A small portion of uric acid is secreted into the gastrointestinal tract for disposal. The secretion of uric acid (via biliary and/or intestine) in the gastrointestinal tract is thought to be by both passive permeation (depending on the urate concentration) and active transportation (mediated by high-capacity urate efflux transporters) (16). The intestinal urate transporters have not been well investigated so far, but recent studies report a high expression level of *ABCG2* gene (encoding a high-capacity urate efflux transporter) in the intestinal epithelium and suggest that the ABCG2 transporter plays an important role in the intestinal secretion of uric acid (17). The secreted uric acid is further degraded by intestine bacteria (18). Especially in the lower intestine tract, uric acid is exposed to a large number of bacteria, e.g., *Escherichia coli*, *Aerobacter aerogenes* and *Paracolobactrum* species. Uricase from these bacteria can catabolise uric acid into carbon dioxide (CO_2) and ammonia (NH_3), which are then reabsorbed or eliminated as intestinal air (19). This breakdown process of uric acid in the gut lumen is known as intestinal uricolysis and is responsible for 30% of the total uric acid disposal (20). Nearly all uric acid

secreted in the gut is completely degraded by the intestinal flora, with a only small amount being found in human faeces (15).

- ***Renal excretion***

It is estimated that renal excretion is responsible for the remaining 70% of the daily uric acid disposal (21). The excretion of urate via the kidney largely depends on renal function. Glomerular filtration rate (GFR) describes the flow rate of filtered fluid through the kidney, which is an indication of the kidney condition. Fractional extraction of urate (FEUR) is used to represent the percentage of filtered urate that is finally excreted via urine. It is calculated as the ratio of urate clearance (C_{UA}) to creatinine clearance (C_{Cr}, an approximation to GFR), with the formula: FEUA= [(U_{UA} × Urine sample volume)/P_{UA}] / [C_{Cr} = (U_{Cr} × Urine sample volume)/P_{Cr}] ×100% = (U_{UA} × P_{Cr})/ (U_{Cr} × P_{UA}) ×100% (U_{UA}: urinary urate concentration; P_{UA}: plasma urate concentration; U_{Cr}: urinary creatinine concentration; P_{Cr}: plasma creatinine concentration).

It is reported that nearly all circulating urate (>95%) is readily available to be filtered at the glomerulus, however, the FEUA in healthy adults is only 10% (range: 7-12%), indicating that the net tubular reabsorption of the filtered urate is about 90% (22). The renal handling process on urate is predominately explained by a classical model including four distinct components: (i) glomerular filtration, (ii) reabsorption in the proximal tubule, (iii) secretion near the terminus of the proximal tubule, and (iv) post-secretory reabsorption near these secretory sites. Specifically, almost all urate is filtered at the glomeruli; subsequently, pre-secretory reabsorption returns the majority of the filtered urate into the early proximal tubule; in the proximal tubule, 50% of the filtered urate is secreted back into the tubular lumen, and then the secreted urate undergoes post-secretory reabsorption resulting in 7-12% of filtered urate load being excreted by the kidney (22).

- ***Molecular mechanisms of urate transport***

The molecular mechanisms underlying the renal and intestinal handling processes of urate are not completely understood, but urate transporters are believed to play pivotal roles. The urate-organic exchanger and voltage-sensitive pathways are suggested as the major two modes of urate transport. Among these identified transporters, URAT1 (organic anion transporter 1) and GLUT9 (glucose transporter 9) are believed to play important roles in renal reabsorption of urate, while ABCG2 (ATP-binding cassette sub-family G member 2) and ABCC4 (ATP-binding cassette sub-family C member 4) are responsible for urate secretion (21).

URAT1, encoded by the *SCL22A12* gene, is the major organic anion transporter (OAT) for urate (23). *SCL22A12* is uniquely expressed in the luminal membrane of the proximal renal tubular epithelium. Urate transport mediated by URAT1 is independent of the sodium-urate co-transport but could be interfered with by organic anions, such as lactate, nicotinate, hydroxybutyrate, acetoacetate, and succinate. Intracellular accumulation of organic anions, which have an affinity with URAT1, facilitates the reabsorption of urate to exchange these anions out of the cell to maintain electronic balance. After being absorbed into the cell, urate then moves across the basolateral membrane into the blood by means of other organic anion exchangers, of which the most important one is GLUT9. GLUT9 is encoded by *SLC2A9* gene and highly expressed in the kidney and weakly expressed in the liver and intestine (24). Apart from its strong ability in urate transport, it has also been previously identified as a fructose transporter (25).

Uric acid secretion seems to be primarily mediated by the voltage-sensitive urate transporter, ABCG2, which substantially localises on the apical side of proximal tubular cells (26). *ABCG2* is also expressed in the intestinal epithelium to a lesser extent, which is believed to contribute to the movement of urate into the gut (17). Another candidate, ABCC4, also known as the multidrug resistance-associated protein 4 (MRP4) or multi-specific organic anion transporter B (MOAT-B), is a member of the adenosine triphosphate-binding cassette transporter family (27). ABCC4 is a novel renal apical organic anion efflux transporter, which mediates the secretion of urate and other organic anions, for instance, cGMP and cAMP (28).

Other molecular mechanisms contributing to urate transport in renal tubular cells and intestinal epithelial cells have also been proposed, but their specific roles in the physiology of urate transport are still under investigation. The identified urate transporters with their localisations and physiological functions are summarised in **Table 1-1** and **Figure 1-3**.

Table 1 - 1: Urate transporters and their functions.

UA transporters	**Coding genes**	**Function**	**Localisations**	**Contribution**
URAT1	*SLC22A12*	Urate-anion exchanger	Apical	Reabsorption
GLUT9	*SLC2A9*	Urate uniporter	Basolateral	Reabsorption
ABCG2	*ABCG2*	Urate uniporter	Apical	Secretion
NPT1	*SLC17A1*	Na-phosphate cotransporter; Organic ion uniporter	Apical	Secretion
NPT4	*SLC17A3*	Organic ion uniporter	Apical	Secretion
OAT4	*SLC22A11*	Organic anion-dicarboxylate exchanger	Uncertain	Reabsorption
OAT1	*SLC22A6*	Organic anion-urate exchanger	Basolateral	Secretion

OAT3	*SLC22A8*	Organic anion-urate exchanger	Basolateral	Secretion
MCT9	*SLC16A9*	Monocarboxylic anion transporter	Uncertain	Uncertain
ABCC4	*ABCC4*	ATP-dependent anion pump; Urate ion pump	Uncertain	Secretion

Source: adapted from reference (21) with permission from the Springer Nature Publisher [reference number:4353060751528].

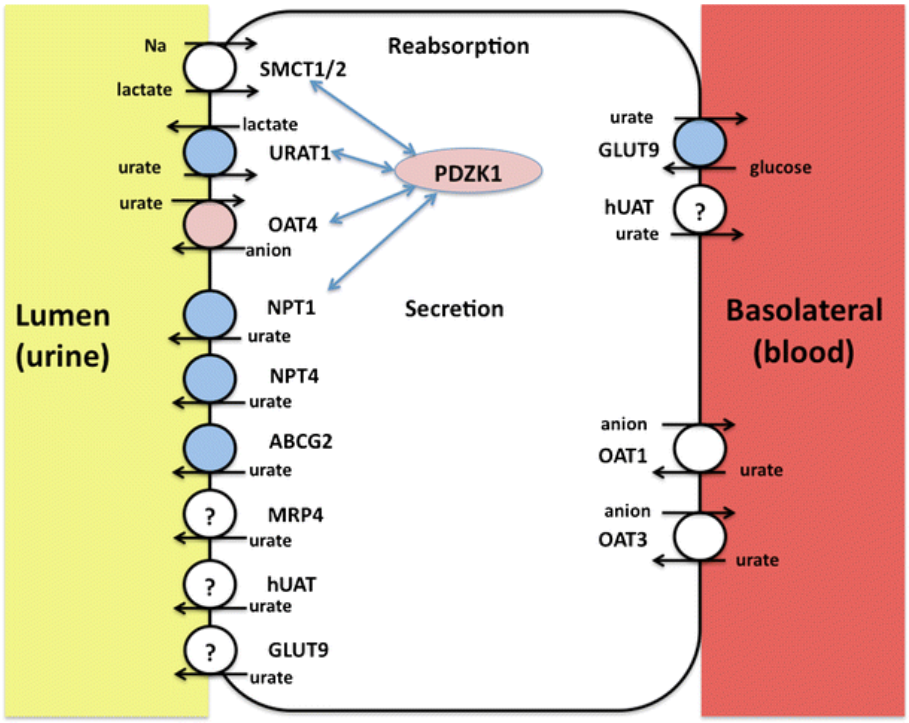


Figure 1 - 3: Models for urate transport.

Blue circles represent function confirmed by genome-wide association studies (GWAS), mutations in humans, and in vitro studies. Pink circles represent function confirmed by GWAS and in vitro studies. White circles represent strong in vitro data. Question marks refer to proposed but uncertain function. OAT, organic anion transporter; (Source: adapted from reference (21) with permission from the Springer Nature Publisher [reference number:4353060751528]).

1.1.3 Biological function

Uric acid has long been regarded as a metabolically inert waste product produced from catalytic activity of xanthine oxidoreductase. However, as increasing evidence shows uric acid has multiple biological roles (behaving as both anti-oxidant and pro-oxidant) in affecting the oxidative status in cultured cells, there are great controversies regarding its role in pathophysiology, with some researchers suggesting that it is simply a marker of xanthine oxidoreductase activity (29), others stating that it may exert beneficial effects as an anti-oxidant (30, 31), and others suggesting that it may induce the risk of oxidative stress due to its pro-

oxidative properties (32, 33). This section is a brief review on the role of uric acid as an anti-oxidant, pro-oxidant or surrogate marker under different conditions.

1.1.3.1 Uric acid as an anti-oxidant

According to an evolutionary hypothesis, the silencing of the uricase genes, resulting in an increased SUA level, provides a survival advantage for ancestors of *Homo sapiens* (34). Uric acid was believed to be one of the most important antioxidants circulating in the blood to protect cells from oxidative damage, thereby contributing to a prolonged lifespan and reduced risk of cancer (35). This hypothesis was based on in vitro experiments which demonstrated that uric acid can scavenge oxygen radicals and thus protect cells from oxidation.

This property of uric acid, as a powerful scavenger of single oxygen peroxyl and hydroxyl radicals (-OH), was firstly demonstrated by Kellogg *et al* (36) and was further characterised by Ames *et al* (35). It has been indicated that uric acid acts as an antioxidant by reacting with a variety of oxidants, such as hydroxyl radical, peroxynitrite, and nitric oxide (37). The most classical scavenging process is the reaction of uric acid with the hydroxyl radical (OH), resulting in the formation of allantoin. In addition, uric acid could also react with peroxynitrite, a vital biological oxidant generated from the reaction of nitric oxide (NO) with a superoxide anion, to convert it into triuret (38, 39). Moreover, uric acid has also been reported to react with NO to produce a nitrosated compound with the capability of donating NO (40). Uric acid can also block ferrous (Fe^{2+}) catalysed oxidation reactions in human (41). However, the ability of uric acid as an antioxidant is limited by several conditions. Specifically, uric acid is unable to scavenge superoxide (O_2^-), and the scavenging process of peroxynitrite requires the presence of ascorbic acid and thiols (32, 38). The antioxidant property of uric acid could also be disabled by some compounds present in the body fluids. For example, the presence of bicarbonate can significantly inhibit the ability of uric acid to prevent the process of tyrosine nitrosylation, which is a critical step of oxidative damage of cellular proteins (42). Additionally, uric acid cannot scavenge lipophilic radicals and has no ability to break the radical chain propagation within lipid membranes (43). These physicochemical findings indicate that the antioxidant effect of uric acid is limited and probably manifested only in the hydrophilic environment of biological fluids, such as plasma.

1.1.3.2 Uric acid as a pro-oxidant

The ability of uric acid acting as a pro-oxidant has always been related to the induction of oxidative stress. Oxidative stress is a condition of excessive production of free radicals and/or

reactive oxygen species (ROS), as well as reduced anti-oxidative ability, which is usually due to the decreased intake or excessive consumption of antioxidants (44). It has been shown that the antioxidant reactions of uric acid with oxidants are accompanied by the formation of a variety of free radicals (45, 46). Radicals derived from these antioxidant reactions represent different degradation degrees of the uric acid molecule, varying from the urate anion (the radical site is located on the five-membered ring structure of uric acid) to carbon-based radicals, such as aminocarbonyl (which is generated after the breakdown of the five-membered ring structure due to the $ONOO^-$ attack) (45, 46). Uric acid itself and/or its downstream radicals seem to induce lipid oxidation, which can stimulate oxidant production in adipocytes (33). The increased level of oxidised lipids could in turn propagate a radical chain reaction and cause oxidative stress. The pro-oxidative effects of uric acid are likely mechanised through activating a nicotinamide adenine dinucleotide phosphate (NADPH) oxidase-dependent pathway, which results in increased production of reactive oxygen species (ROSs) (33). In contrast to its anti-oxidative effect, the pro-oxidative property of uric acid inducing oxidative stress mainly occurs in the intracellular environment, particularly in adipocytes (32).

1.1.3.3 Uric acid as a surrogate marker of XOR activity

Uric acid has also been recognised as a surrogate marker of xanthine oxidoreductase (XOR) activity. XOR is an enzyme that catalyses the production of uric acid with two forms, xanthine dehydrogenase and xanthine oxidase (47). Xanthine dehydrogenase has an affinity for oxidised NAD to convert it into uric acid and NADH (48). Xanthine oxidase first catalyses the oxidation of hypoxanthine to xanthine and further catalyses the oxidation of xanthine to uric acid, with the production of two oxidant molecules (one superoxide is produced in each step) (47). The ROSs generated by XOR activity from the formation of uric acid are thought to play a vital role in increasing oxidative stress. Hence, in this situation it becomes unclear if uric acid itself induces the damaging oxidative stress or it simply represents a surrogate marker of high-level oxidative stress due to the increased XOR activity. Although many experimental studies in animals support the pathophysiologic role of uric acid as discussed above (37), evidence from human studies is unclear. Clinical trials have been performed to examine the role of uric acid in disease states by utilising XOR inhibitors (49, 50). Xanthine oxidase inhibitors can reduce uric acid formation, but they can also decrease XOR induced superoxide generation, thus it is uncertain if any benefits observed with xanthine oxidase inhibitors are due to the decreased SUA level or due to the reduction of xanthine oxidase-associated oxidants. Given these, more efforts are required to clarify the pathophysiologic role of uric acid in humans.

1.2 Hyperuricaemia

Abnormally elevated SUA level is the most common dysregulation of urate metabolism. Hyperuricaemia is a clinical term traditionally applied to the settings in which SUA levels are elevated with or without any clinical manifestation. Two-thirds or more of individuals with hyperuricaemia may have no clinical symptoms (asymptomatic hyperuricaemia), while one-third have the signs of monosodium urate (MSU) crystal deposition in joints (e.g., gout) or kidneys (e.g., nephrolithiasis) (51). Increased uric acid production, decreased uric acid excretion, or a combination of both problematic processes are the commonest causes of hyperuricaemia. The decreased efficiency of urate excretion is responsible for about 85%-90% of hyperuricaemia, while the remaining 10%-15% is caused by uric acid overproduction, which is often induced by genetic defects, disease conditions or intake of drugs or purine-rich diet (52). The definition/classification, aetiology, epidemiology and management of hyperuricaemia are presented in this section.

1.2.1 Definition of hyperuricaemia

There is no universally applicable definition for hyperuricaemia. It could be defined in several ways (**Table 1-2**), including population values, physiochemical cut-point, or the levels associated with disease risk.

A statistically based definition for hyperuricaemia refers to the SUA level more than two standard deviations (SDs) above the mean of the healthy population. However, given the non-normal distribution of SUA level in most populations and variations based on ethnicity, age and sex, the statistical definition of hyperuricaemia varies among populations (53).

A physiochemical definition of hyperuricaemia based on the solubility limit of uric acid in plasma is preferred when compared to the statistical definition (53). This physiochemical definition corresponds to a SUA level of more than 416 μmol/L (7.0 mg/dL) as measured by an automated enzymatic method or corresponds to a SUA level of more than 476 μmol/L (8.0 mg/dL) as measured by a colorimetric method.

Epidemiological studies have shown that associations of high SUA level with non-crystal deposition disorders are always observed without reaching the saturating concentration. In this case, it is indicated that the physiochemical definition may not be appropriate for clinical practice (54). For practical and/or clinical purposes, an alternative definition of hyperuricaemia is suggested by some experts, in which a concentration of SUA exceeding 357 μmol/L in blood is used as the threshold of clinically relevant hyperuricaemia (53, 54). This reference value

integrates the estimated threshold for the lifelong risk for clinical consequences of hyperuricaemia and fits the recommended goal (<357 μmol/L) of clinical urate-lowering treatment in gout patients (55). Additionally, another significant threshold of SUA level is >476 μmol/L and this reference value is applied as the threshold of initiating evaluation and lifestyle/pharmacologic intervention on patients with asymptomatic hyperuricaemia (55).

Table 1 - 2: Definitions for hyperuricaemia.

Definition	Criteria	Weakness
Statistical definition	SUA level >2 standard deviations above the mean value of healthy population	Variations due to ethnicity, age, and sex, *etc*.
Physiochemical definition	Based on the saturation point of uric acid when monosodium urate crystals occur.	Changes according to temperature and the PH.
Definition based on the treatment target of gout and the potential of other diseases	For all gout patients, the treatment target of SUA level is <357 μmol/L; for patients with severe gout, the treatment target of SUA level is <297 μmol/L.	Differs between conditions.

Source: adapted from reference (53, 56) with permission from the Elsevier publisher [reference number: 4353120627211].

Among these definitions, SUA level above 416 μmol/L is the most widely used threshold to define hyperuricaemia, as there appears to be little disagreement regarding to the physiochemical characteristics of uric acid. Hyperuricaemia could be further classified into two categories based on the causes: (i) primary hyperuricaemia refers to a rise of SUA level due to genetic deficiencies; (ii) secondary hyperuricaemia refers to excessive uric acid production or decreased renal clearance caused by acquired clinical disorders, drugs, purine-rich diet, or toxins. Primary hyperuricaemia usually occurs in childhood and last indefinitely, while secondary hyperuricaemia can occur at any age.

1.2.2 Aetiology

The underlying aetiological mechanisms of hyperuricaemia could be classified into three categories: increased uric acid production (e.g., inherited enzyme defects, increased cell turnover and tissue hypoxia, purine-rich diet), decreased uric acid excretion (e.g., kidney disease, certain drugs, metabolic or endocrine diseases), and mixed type (e.g., high levels of alcohol and/or fructose, moderate/severe exercise, starvation). The main causes of hyperuricaemia are described below.

1.2.2.1 Inherited enzyme defects

Genetic mechanisms inducing hyperuricaemia include overproduction of uric acid due to hypoxanthine–guanine phosphoribosyltransferase (HGPRT) deficiency or, phosphoribosylpyrophosphate (PRPP) synthetase overactivity or glucose-6-phosphatase (G6PT) deficiency (57). These inherited defects often lead to early development of severe hyperuricaemia.

- ***Hypoxanthine-guanine phosphoribosyltransferase (HGRPT) deficiency***

HGRPT is a transferase that catalyses the conversion of hypoxanthine to inosine monophosphate and guanine to guanosine monophosphate. HGRPT deficiency is a rare inherited disorder caused by mutations in the *HGPRT* gene localised on the long arm of the X chromosome (57). Absence of HGRPT prevents the normal metabolism of hypoxanthine resulting in excessive uric acid production. A complete loss of HGPRT activity results in a severe disorder, called Lesch-Nyhan syndrome(58), and a partial deficiency in HGPRT activity causes a moderate disorder, called Kelley–Seegmiller syndroms (59). Both of them are characterised by overproduction of uric acid.

- ***Phosphoribosylpyrophosphate (PRPP) synthetase over-activity***

Phosphoribosylpyrophosphate (PRPP) synthetase over-activity is an X-linked disorder of purine metabolism (57). Mutations in the *PRPP synthetase (PRPS)* gene lead to over-activity of PRPP synthetase and therefore increase the production of PRPP (60). Increased level of PRPP enhances *de novo* synthesis of purine nucleotides. When excessive production of purine nucleotides exceeds the need of the human body, the excess purine nucleotides are catabolised leading to overproduction and accumulation of uric acid. Clinical consequences of PRPP synthetase over-activity include an early-onset severe form, characterised by gout, urolithiasis, and neurodevelopmental anomalies (severe PRPP synthetase over-activity), and a mild late-onset form with no neurologic damage (mild PRPP synthetase over-activity). Both of them are characterised by overproduction of uric acid.

- ***Glucose-6-phosphatase (G6PT) deficiency***

Deficiency in G6PT is inherited as an autosomal recessive trait, leading to type I glycogen storage disease (61). Clinical consequences associated with this defect are increased uric acid production and symptoms of gout (61). The inability to dephosphorylate G6PT leads to an increased diversion of glucose into the pentose phosphate pathway (PPP). One of the main

products of PPP is ribose 5-phosphate. Increased production of ribose 5-phosphate leads to substrate-level activation of PRPP synthetase. Over-activity of PRPP synthetase has the same consequence as the inherited defect in the *PRPS* gene resulting in overproduction of uric acid as explained above.

1.2.2.2 Accelerated cell turnover and tissue hypoxia

Uric acid overproduction also occurs in the setting of increased cell turnover or tissue hypoxia. Certain types of tumours, such as metastatic cancer, multiple myeloma, leukaemia and lymphomas, can induce hyperuricaemia. Overproduction of uric acid in these cancers results both from the increased turnover rate of cell death and from the use of chemotherapy agents. With rapidly growing tumours, there are high rates of cell turnover and tumour proliferation, generating a large amount of purine nucleotides. Chemotherapy treatments can also cause overproduction of uric acid due to tumour lysis syndrome (TLS), characterised by a rapid amount of cellular destruction (62). Massive cell death and nuclear breakdown generates large quantities of nucleic acids. Of these, the purine nucleotides are catabolised into uric acid via the purine degradation pathway. At a high level of uric acid production, hyperuricaemia and MSU crystals occur. Additionally, overproduction of uric acid could also be induced by tissue hypoxia, under which ATP is consumed with the generation of its isoform xanthine oxidase (a necessary enzyme catalysing the formation of uric acid) (63). This may probably explain the high SUA level in individuals with congestive heart failure, congenital cyanotic heart disease, high altitude hypoxia, or obstructive sleep apnoea due to hypoxia (63, 64).

1.2.2.3 Renal insufficiency

Decreased efficiency of renal excretion of uric acid is responsible for about 85%-90% of hyperuricaemia, due to the complexity of the renal handling of uric acid, the sensitivity of kidney to metabolites and drugs, and mutations in urate transporter coding genes. Under-excretion of urate appears to be a combination of decreased glomerular filtration, decreased tubular secretion, and increased tubular reabsorption. With acute and/or chronic kidney diseases, the kidney loses its ability to filter uric acid properly, leading to accumulation of uric acid in the blood. Decreased tubular secretion of urate often occurs in patients with metabolic or endocrine diseases (e.g., diabetic/starvation ketoacidosis, lactic acidosis, and ethanol or salicylate intoxication), as organic acids (e.g., lactate, acetoacetate, and beta-hydroxybutyrate) accumulate in these conditions and compete with uric acid for tubular secretion. Additionally, diuretic therapy can induce hyperuricaemia by enhancing reabsorption of uric acid in the distal tubule.

Some medications, such as pyrazinamide, salicylates, nicotinic acid, ethambutol, cyclosporine and cytotoxic agents, can interfere with the renal urate excretion, causing a rise in SUA level.

1.2.2.4 Diet, starvation, and exercise

Hyperuricaemia may also result from diet rich in purines, alcohol or fructose. Diet containing a high concentration of purines includes organ meat such as kidney and liver, red meat, poultry, fish, sardines, anchovies, mushrooms, yeast and beer (65). Exogenous intake of purines from diet can enrich the purine pool and enhance the degradation of purine nucleotides with the formation of excessive uric acid. Alcohol consumption can increase uric acid synthesis due to enhanced turnover of ATP from the conversion of acetate to acetyl-CoA during the metabolism of ethanol (66). In addition, acute large amounts of alcohol consumption can increase the production of lactate that can reduce renal urate excretion and exacerbate hyperuricaemia (67). Fructose, a simple sugar widely present in sucrose, honey, and fruits, can also rapidly increase SUA level. This is in part due to its rapid phosphorylation in hepatocytes with ATP consumption, intracellular phosphate depletion, and the stimulation of AMP deaminase with the generation of uric acid (68). Fructose intake also stimulates the synthesis of uric acid from amino acid precursors (68). The remarkable increase in fructose intake in populations may play a role in the rising level of SUA worldwide (69).

Starvation can also increase SUA level by both increasing uric acid production and decreasing uric acid excretion. In starvation, the body breaks down its own (purine-rich) tissues for energy, thus rapid weight loss exacerbates hyperuricaemia (70, 71). Starvation can also impair the ability of the kidney to excrete uric acid, because of the competition between uric acid and ketones for transport (72).

In addition, SUA level may be affected by exercise, with moderate exercise reducing SUA level (probably because of the increased renal blood flow) and severe exercise increasing SUA level (probably due to the ATP consumption and thereby the formation of adenosine and xanthine) (73).

1.2.3 Epidemiology of hyperuricaemia

Elevation of SUA level is evident in all world regions, particularly in Asia and the North America, where epidemiological data are abundant. However, the global burden of hyperuricaemia is difficult to be precisely quantified due to considerable regional variations. The purpose of this section is to summarise the epidemiologic characteristics of hyperuricaemia in terms of prevalence and incidence in different world regions. In order to capture the global or

national data on the epidemiology of hyperuricaemia, a systematic literature search was conducted in the MEDLINE and EMBASE on 12th November 2017 by using the MeSH terms: "gout or hyperuri*" AND "prevalen* or inciden* or epidemiolo*". The retrieved literature was screened in title, abstract and full text review by myself. Population-based epidemiological studies on hyperuricaemia were eligible for inclusion and their findings are described below.

1.2.3.1 Geographical prevalence

In this literature review, 34 articles were finally included, reporting the population-based prevalence/incidence of hyperuricaemia for a total of 24 countries, which are grouped into 6 WHO regions (**Table 1-3**). The proportion of individuals with hyperuricaemia was highly variable (range: 4.9%-53.8%) across various regions of the world with data lacking for many countries.

- ***Western Pacific region***

The highest prevalence of hyperuricaemia was reported in the Western Pacific region within the ethnic group of Taiwanese aboriginals, for which the estimate was 53.8% (74). The prevalence of hyperuricaemia in mainland China was between 11.0%-32.1% from the population-based surveys of several areas (75-79). Data from Japan reported a high prevalence of 34.5% (80, 81). In contrast, a very low prevalence of 4.9% was reported in South Korea (82). In Oceania area, the Pacific islanders also had a high prevalence of hyperuricaemia (83-85). In the Australia Nation Health Survey, the crude prevalence estimate of hyperuricaemia in the general population was 23.0% (86), while the corresponding estimate for the white Australians was 12.0% (87). In New Zealand, hyperuricaemia was found in 15.3% of non-Maori adults (88), while the corresponding estimate for Maori individuals was in the range of 27.8%-49% (88, 89).

- ***South-East Asia***

Data from South-East Asia were only available for two countries, in which Indonesia reported a high prevalence of 34.5% (80, 81), and the corresponding estimate for Thailand was18.4% (90).

- ***Americas***

Data from the US National Health and Nutrition Examination Survey (NHANES) 2007–2008 study reported a 21.6% prevalence of hyperuricaemia in the US adults (91). The prevalence was much lower in European Americans (7.5%) and Native Americans (range: 3.3%-7.2%) (92, 93).

A sample population in Mexico City was reported to have a hyperuricaemia prevalence of 26.5% in men and 19.8% in women (94). In Brazil, the prevalence of hyperuricaemia was reported in the range of 3.2%-5.6% for different age groups (95, 96).

- ***Eastern Mediterranean***

For the Eastern Mediterranean region, only one study was identified, which reported a prevalence of 8% for Saudi Arabia (97).

- ***Europe***

The prevalence of hyperuricaemia in Europe was reported for 6 countries with a range from 6.6% to 19.0%. In the UK, the prevalence of hyperuricaemia was reported to be 6.6%-8.0% (98, 99). A similar hyperuricaemia prevalence of 6.6% was reported in the Finnish population (100). A regional community-based study in Italian population reported a range of 8.5%-11.9% for hyperuricaemia prevalence (101). Studies in France, Turkey and Russia, reported very similar prevalence of 17.6%, 19.0%, and 16.8%, respectively (102-104).

- ***Africa***

In Africa, the prevalence of hyperuricaemia was reported to be 20.5% in Nigerian men (105), 25.0% in black Africans in Angola (106), 35.2% in Seychellois men and 8.7% in Seychellois women (107).

Table 1 - 3: Prevalence of hyperuricaemia in different world regions.

WHO regions	Population	Prevalence (%)	Reference
Western Pacific region			
China, Mainland	Han Chinese	11.0-32.1%	(75-79)
China, Taiwan	Taiwanese Aborigines	53.8%	(74)
Japan, Okinawa	Japanese	34.5%	(80)
South Korea	Middle-aged men	4.9%	(82)
Australia	White Australian	12.0%	(87)
	General Population	23.0%	(86)
New Zealand	Non-Maori	15.3%	(88)
	Maori	27.8-49.0%	(88, 89)
Cook Islands	Rarotongans	44.0%	(83)
	Pukapukans	48.5%	(83)
Samoa	Samoans (urban)	36.4%	(84)

	Samoans (rural)	43.3%	(84)
Fiji	Fijians (urban)	32.4%	(85)
	Fijians (rural)	16.9%	(85)
South-East Asia			
Indonesia, Java	Malayo-Polynesians	24.3%	(81)
Thailand	Thai	18.4%	(90)
Americas			
United States (USA)	Americans of multiple ancestries	21.2%	(91)
USA, Michigan	European Americans	7.5%	(92)
USA, Montana	Native Americans	3.3-7.2%	(93)
Mexico	Mexican men	26.5%	(94)
	Mexican women	19.8%	(94)
Brazil	Amerindians	3.2-5.6%	(95, 96)
Eastern Mediterranean			
Saudi Arabia	Saudis	8.0%	(97)
Europe			
UK	English	6.6-8.0%	(98, 99)
Finland	Finnish	6.6%	(100)
France	French	17.6%	(102)
Turkey	Turkish	19.0%	(103)
Russia	Russian	16.8%	(104)
Italy	Italian	8.5-11.9%	(101)
Africa			
Nigeria	Nigeria (men)	20.5%	(105)
Angola	Black Africans	25.0%	(106)
Seychelles	Seychellois (men)	35.2%	(107)
	Seychellois (women)	8.7%	(107)

1.2.3.2 Incidence and time trend

Published data on hyperuricaemia incidence are limited. However, it is evident that the trend of hyperuricaemia has been steadily rising in recent decades (108, 109). The rising incidence of gout indirectly reflects the progressive increase in SUA level (110). Over time, the incidence of gout in the US adults increased from 0.03% in 1978 to 0.05% in 1996 (111). Similarly, the

NHANES found that the self-reported lifetime prevalence of gout increased from 2.6% in the NHANES 1988–1994 to 3.8% in the NHANES 2007–2010 (112). Correspondingly, SUA level had also consistently increased over the interval between the two NHANES studies (91). The UK General Practice Research reported that the gout incidence increased steadily from a low of 11.9 cases per 10,000 person-years in 1991 to a high of 18.0 cases per 10,000 person-years in 1994 (113). Subsequently, the Royal College of General Practitioners Weekly Returns Service reported that the annual prevalence of gout in UK increased slightly from 0.43% in 2001 to 0.47% in 2007 (114). Data from successive surveys undertaken in New Zealand showed a remarkable increase of gout prevalence in both European descents and Maori residents (115). The prevalence of gout in China also increased. Successive population surveys in the city of Qingdao found the prevalence of gout was 0.36% in 2002 and had increased to be 0.53% in 2004 (76, 116). Successive surveys conducted in Shantou area reported a prevalence of 0.17% in 1992, 0.15% in 1995 and 0.26% in 1999 (117).

1.2.3.3 Sex-, age-, and ethnicity-related demographics

Hyperuricaemia is far more common in men than in women. Primary hyperuricaemia frequently begins at puberty in men but is usually delayed until after menopause in women. The sex difference is in concordance with the fact that SUA levels in adult men exceed that in women at reproductive age, but after menopause, SUA levels in normal women increase and approximate that of normal men at a corresponding age. Similarly, when under 65 years old, males have a fourfold higher prevalence of gout than females, while this ratio reduces to 3:1 male to female when over 65 years (108). The substantial sex difference probably relates to the enhancement of renal urate clearance caused by oestrogenic compounds.

The mean SUA level in children is lower than that in adults. The upper limit of the normal reference range of SUA level in children is 297 μmol/L, while the corresponding value is 416 μmol/L in adult men and 357 μmol/L in adult women (118, 119). Consistently, the risk of hyperuricaemia also increases with age. As reported in the NHANES survey, the hyperuricaemia prevalence rises with the increasing age groups (91, 108). In particular, the incidence of hyperuricaemia and gout is more evident from the age of 30 in men and after the age of 50 in women (120).

The risk of hyperuricaemia and gout varies across different ethnicities (108). Hyperuricaemia was much more common in the Pacific Maori than Europeans (27.1% *vs.* 9.4% in males, 26.6% *vs.* 10.5% in females) (88). In the US, African Americans develop hyperuricaemia more frequently than European Americans. It was estimated that African American females had 2.3

times higher risk of developing hyperuricaemia comparing to European American females (121). The Hmong population in southern China also suffered higher SUA level (122). Filipinos in the US are at a higher risk of elevated SUA level than other ancestries in the US (123).

1.2.4 Clinical evaluation

Generally, health screening and laboratory evaluation of medical conditions unrelated to urate crystal deposition diseases do not routinely include measurement of SUA level (55, 124), unless in some cases where individuals were identified with SUA level >476 μmol/L, a follow-up test is suggested to confirm the presence of sustained hyperuricaemia (55, 124).

1.2.4.1 Confirmatory test

A repeated measurement of SUA level is needed to confirm the presence of sustained hyperuricaemia. For patients with a SUA level less than 416 μmol/L in the confirmatory test, further follow-up and evaluation are usually not performed. For patients with SUA level between 416-476 μmol/L in the confirmatory test, a repeat test should be performed during the following 6-12 months. For patients with a SUA level greater than 8 mg/dL in the confirmatory test, a more detailed evaluation should be initiated. The decision of no further evaluation on patients with SUA level ⩽476 μmol/L is a practical compromise, considering the low incidence of gout in hyperuricaemia patients with SUA level between 416-476 μmol/L (125).

1.2.4.2 Preliminary evaluation

Evaluation is initiated for subjects with SUA level greater than 476 μmol/L in the confirmatory test. The preliminary evaluation includes a thorough history and physical examination, such as medical conditions, diet or lifestyle habits, pharmaceutical therapies, toxin exposure, or any known familial genetic disorders that may cause hyperuricaemia. Initial laboratory tests include a complete blood count and differential leukocyte counts, urinalysis, measurement of renal function and examinations on chemical profile, including electrolytes, calcium, and liver chemistries. In general, the initial evaluation would identify approximately 80%-90% causes of hyperuricaemia. Otherwise, additional laboratory examinations are considered to detect some specific causative factors. For instance, high level of SUA (>595 μmol/L in children or >714 μmol/L in adolescents) may indicate a sign of underlying disorders, such as lymphoproliferative or myeloproliferative state, or inherited enzyme defects, which are indications for appropriate imaging, pathologic, biochemical, and/or genetic measurements.

In summary, the main aim of preliminary evaluation of hyperuricaemia is to identify:

- Patients at high risk of gout or urolithiasis who need anti-hyperuricaemia treatment.
- Drug or toxin induced hyperuricaemia that can be removed or substituted to relieve the hyperuricaemia status. This is especially important to patients for which lifestyle modifications and/or alternative pharmacologic medications are available.
- Individuals whose hyperuricaemia is a sign of underlying disorders (e.g., inherited enzyme defects, lymphoproliferative or myeloproliferative state).

1.2.4.3 Further evaluation

A measurement of FEUR (fractional extraction of urate), representing the percent of urinary urate excretion per unit of GFR, should be performed for high level of SUA without identifying any cause during their initial evaluation. The FEUR can be determined by measuring the urate and creatinine concentrations in both blood and urine collections (see the *Chapter 1, Section 1.1.2.2 "Renal excretion"* for more explanations on FEUR). The blood and urine collections for these studies should be carried out under the condition in which the individual has a standard diet, without consuming alcohol and drugs known to affect uric acid metabolism. The determination of FEUR will help to distinguish between causes resulting from increased uric acid production with hyperuricosuria (FEUR>10%) and causes resulting from decreased renal clearance (FEUR<6%) (126).

Urinary uric acid excretion greater than 800 mg/day (4.8 mmol/day) or FEUR >10% is defined as hyperuricosuria, indicting excessive uric acid production from either exogenous (dietary) or endogenous purine catabolism (127). The exogenously and endogenously determined uric acid overproduction can be distinguished by the following clinical evaluations. Patients under clinical evaluation should have an isocaloric, purine-reduced diet (consuming 1 gram/kg dairy protein per day without intake of meat, seafood, alcohol and medications affecting uric acid metabolism) for 3-5 days. For patients with UA excretion >670 mg/day (4 mmol/day), endogenous causes of uric acid overproduction (e.g., inherited enzyme defects, or disorders resulting in increased rate of cell turnover) should be considered. Otherwise, excessive dietary purine consumption is confirmed as the cause of hyperuricosuria for patients whose uric acid excretion (less than 670 mg/day [4 mmol/day]) decreases to normal values on a purine-reduced diet. Under-excretion of urate but with normal renal function is indicative of genetic defects in urate transporter coding genes (functioning either as reduced secretion or as enhanced reabsorption), or secondary to reduced renal perfusion (e.g., diuretics). The distinction can provide further guidance on the choice of anti-hyperuricaemia drugs for individuals needing pharmacologic treatment.

1.2.5 Treatment

Despite the high prevalence of hyperuricaemia, pharmacologic therapies on this biochemical aberration are not widely recommended (55, 124). The implementation of either non-pharmacologic (lifestyle-based) intervention or pharmacologic therapies should be determined based on the estimates of the clinical consequences of hyperuricaemia. Additionally, for pharmacologic urate-lowering therapy, the estimated risk of clinical consequences related to hyperuricaemia should be weighed against the potential benefits and risks, given urate-lowering medications (e.g., allopurinol or colchicine) can induce rare but very severe and even life-threatening adverse reactions.

1.2.5.1 General principles

- ***Asymptomatic hyperuricaemia***

Patients with hyperuricaemia (SUA level >476 μmol/L) but without evidence of urate crystal deposition should be offered a plan for non-pharmacologic (lifestyle) management to reduce SUA level. The lifestyle interventions include adjustment of dietary composition, reduction of alcohol intake, control of body weight, and regular moderate exercise (124, 128). Consumption of dairy products, particularly low-fat dairy products or some dietary supplementations such as with vitamin C, was found to be associated with a substantially reduction in SUA level (129).

Attentions should also be taken in the management of some accompanying diseases. Specifically, the use of medications that may increase SUA level should be avoided for hyperuricaemia patients; on the contrary, the use of medications that can reduce SUA level and/or decrease the risk of gout is much preferred. For example, the use of fenofibrate for hyperlipidaemia, and losartan or calcium channel blockers for hypertension should be promoted among hyperuricaemia patients (130, 131). In contrast, antihypertensive drugs, such as angiotensin-converting enzyme (ACE) inhibitors, thiazide or loop diuretics, non-losartan angiotensin II receptor blockers, and beta blockers, that may raise SUA level, should be avoided when possible (130).

- ***Hyperuricaemia with gout and urate nephrolithiasis***

For subjects whose asymptomatic hyperuricaemia is accompanied by MSU crystal deposition demonstrated only on imaging but with no occurrence of gout, the same approach as for other asymptomatic individuals with hyperuricaemia should be taken. However, for hyperuricaemia patients with gouty arthritis, appropriate urate-lowering pharmacotherapy should be started with

a recommended goal range of SUA level <357 μmol/L (55). Similarly, urate-lowering pharmacotherapy for the prevention of urate stone is not warranted in most individuals, unless a urate stone is discovered. The preferred treatment for hyperuricaemic individuals in the presence of urinary urate stone is hydration (fluid intake >2 L/day) and urinary alkalinisation with administration of potassium bicarbonate or potassium citrate, instead of allopurinol (132).

- ***Sustained high levels of hyperuricaemia with hyperuricosuria***

Pharmacological urate-lowering therapy is considered for individuals with sustained marked hyperuricaemia or a less marked degree of hyperuricaemia but with sustained hyperuricosuria who are at high risk of acute uric acid nephropathy or recurrent bouts of acute renal failure. Sustained high levels of hyperuricaemia and hyperuricosuria always occur among individuals with purine and/or uric acid overproduction due to inherited enzyme defects in purine and/or ATP metabolism or genetic polymorphisms of urate transporter genes resulting in reduction of renal uric acid clearance, or due to clinical disorders associated with accelerated cell turnover. Under these circumstances, even when hyperuricaemia is not presented with any urate-induced disorders (i.e., gout, urate nephrolithiasis), a urate-lowering goal of SUA level <476 μmol/L is recommended to prevent acute renal failure (133).

1.2.5.2 Urate-lowering medications

There are several types of urate-lowering drugs available (134, 135), including (i) xanthine oxidase inhibitors (XOIs): allopurinol and febuxostat; (ii) uricosuric agents: probenecid, benzbromarone, and lesinurad; (iii) uricase: pegloticase and rasburicase

- ***Xanthine oxidase inhibitors (XOIs)***

XOIs are likely to be effective in all circumstances for urate-lowering therapies. The XOI allopurinol is used as first-line anti-hyperuricaemia therapy (134). However, some individuals showed adverse reactions, limiting the use of XOIs (136). Allopurinol and febuxostat are the available forms of XOIs. Oxypurinol is an active metabolite of allopurinol that was previously available as a substitute for patients intolerant to allopurinol.

Allopurinol — The urate-lowering effect of allopurinol is due to its inhibition of xanthine oxidase (xanthine dehydrogenase) activity (137). On one hand, allopurinol, along with oxypurinol, produces inactivation of xanthine dehydrogenase, resulting in decreased UA formation. On the other hand, it can reduce the total urinary excretion of purines due to its inhibition on purine synthesis via drug-derived enhancement of purine base reutilisation and

reduction of purine catabolism. Although allopurinol is very effective in urate-lowering therapy, side effects and adverse reactions occur occasionally. It is estimated that 3%-5% of treated patients would experience rash, diarrhoea, drug fever, leukopenia or thrombocytopenia (136, 138). Sometimes, a potentially fatal adverse event, allopurinol hypersensitivity syndrome (AHS), consisting of erythematous rash, drug fever, hepatitis, eosinophilia, and acute renal failure, may occur in <0.1% of treated patients (136, 138).

Febuxostat — Febuxostat is another type of XOI. Unlike allopurinol, which is a purine base analogue, febuxostat is a thiazolecarboxylic acid derivative that inhibits xanthine oxidase by occupying a channel in the xanthine oxidase dimer (139). Febuxostat produces a dose-dependent effect in the reduction of SUA level (140). Compared to allopurinol, febuxostat is superior in urate-lowering efficacy, but has greater incidence of nausea, arthralgia, rash and abnormalities in liver function test. Thus, hepatic transaminase enzyme levels should be monitored regularly among febuxostat-treated patients (141, 142).

- ***Uricosuric drugs***

Uricosuric drugs are weak organic acids that enhance the renal clearance of urate by inhibiting the urate reabsorption mediated by urate-anion exchangers in the kidney (23). Uricosuric agents include probenecid, sulfinpyrazone, benzbromarone, and lesinurad. Other drugs with modest uricosuric effect include the anti-hypertension drug, angiotensin II receptor antagonist losartan, and anti-hyperlipidaemia drug, fenofibrate (143). Hyperuricaemic patients with renal under-excretion of uric acid are candidates for uricosuric drugs (144). Probenecid and sulfinpyrazone are effective for most patients but ineffective for those with impaired renal function (134). Benzbromarone is likely more effective for patients with mild/moderate renal insufficiency (creatinine clearance 30-59 mL/min), and patients who are not tolerant for allopurinol/febuxostat therapy (134). Lesinurad is an inhibitor of urate transporters (URAT1 and OAT4), involving in uric acid reabsorption in the kidney. Lesinurad should only be used in combination with XOI in patients who have not achieved target SUA level with an XOI alone (145). The major side effects of uricosuric drugs include rash, acute gouty arthritis, gastrointestinal intolerance, and uric acid stone formation (146).

- ***Uricase***

Uricase (urate oxidase) is an enzyme that catalyses oxidation of uric acid into the more water-soluble compound, allantoin. The important criterion in support of uricase treatment is the requirement for reduction or reversal of gout symptoms, particularly for patients with severe

gout on whom treatment with other urate-lowering agents is not effective (134). Pegloticase and rasburicase are two forms of widely used uricase. Pegloticase is a porcine uricase modified by covalent linkage to polyethylene glycol. Pegloticase is effective in the relief of acute gout attacks, but it can lose the urate-lowering effectiveness due to the effects of pegloticase antibodies (147). Rasburicase is developed from *Aspergillus flavus* to minimise the risk of contaminant-related allergic reactions and is widely used to prevent acute urate nephropathy due to TLS in patients with lymphoma and leukaemia (148).

1.3 Genetic polymorphisms

The heritability of SUA levels is estimated to be 40-70%, indicating the importance of its genetic determinants (149). In 2009, a GWAS on serum urate, including 28,141 individuals of European descent, was performed by the European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium and identified nine independent genetic loci (150). Subsequently, in a meta-analysis of 48 genome-wide scans including 110,347 individuals of European descent, the Global Urate Genetics Consortium (GUGC) identified 28 genetic loci (151). When incorporating the GWAS findings of the ENGAGE and GUGC consortia, a total of 31 genetic risk loci were identified in relation to serum urate level, including *SLC2A9, ABCG2, SLC17A1, GCKR, SLC22A11, SLC22A12, PDZK1, SLC16A9, LRRC16A, INHBC, RREB1, HNF4G, SFMBT1, TRIM46, OVOL1, IGF1R, VEGFA, A1CF, BAZ1B, UBE2Q2, ATXN2, TMEM171, HLF, BCAS3, ORC4L, INHBB, NFAT5, STC1, PRKAG2, MAF,* and *PRPSAP1*. Despite the success of the GWAS approach in describing the genetic background of serum urate, a detailed understanding of gene functions in the regulation of SUA level is still lacking, with the exception of urate transporter coding genes. To further describe the nature of these genetic risk loci, the functions of genes mapped by the consortia were annotated by using the GeneCards (http://www.genecards.org/) database. GeneCards is an integrative web-based human gene database that integrates gene-centric data from ~125 web sources and provides comprehensive information on all annotated and predicted human genes.

1.3.1 Urate transporter loci

Many of the identified loci from GWAS harbour genes encoding urate transporters or involving molecular pathways contributing to urate transport. Gene functions for the urate transporter loci are displayed in **Table 1-4**.

Table 1 - 4: Gene functions for the urate transporter loci identified in SUA GWAS.

SNP	CHR	GENE	Gene function (from http://www.genecards.org/)
rs478607	11	*SLC22A12 (Solute Carrier Family 22 Member 12)*	• The *SLC22A12* gene encodes the member 12 of organic anion transporter (OAT1) family which was the first characterised urate transporter regulating serum urate level (23). This protein is an integral membrane protein primarily localised in epithelial cells of the proximal tubule of the kidney and mutations in this gene cause renal hypouricaemia (152). • Its related biological pathways involve the transport of glucose and other sugars, urate, bile salts and organic acids, metal ions and amine compounds (153). • GO annotations related to this gene include the PDZ domain binding and urate transmembrane transporter activity. Its biological function is a pharmacologic target for urate-lowering therapy of uricosuric agents. (153) • Diseases associated with *SLC22A12* include renal hyperuricaemia and hypouricaemia (153).
rs12498742	4	*SLC2A9 (Solute Carrier Family 2 Member 9)*	• The *SLC2A9* gene encodes the member 9 of *SLC2A* facilitative glucose transporter family (*GLUT9*), characterised as a urate transporter and involved in renal urate re-absorption (25, 154).Members of this family play an important role in maintaining glucose homeostasis (155). • Its related pathways involve the transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds (156). • GO annotation related to this gene includes glucose transmembrane transporter activity (156). • Disease associated with the *SLC2A9* gene include renal hyperuricaemia and hypouricaemia (156).

SNP	CHR	GENE	Gene function (from http://www.genecards.org/)
rs2231142	4	*ABCG2 (ATP Binding Cassette Subfamily G Member 2)*	• The *ABCG2* gene encodes the member 2 of ATP Binding Cassette Subfamily (ABCG2) (157). The superfamily of ATP-binding cassette (ABC) transporters transports various molecules across extra- and intra-cellular membranes (157). • Transporter ABCG2 has been shown to act as a urate transporter contributing to the excretion of urate in the kidney. This protein also functions as a xenobiotic transporter which may contribute to multi-drug resistance (158). • GO annotations related to this gene include ATPase activity and the activity of protein homo-dimerisation (159). Its biological function is the transport of urate, glucose and other sugars, bile salts and organic acids, metal ions and amine compounds (159). • Diseases associated with the *ABCG2* include hyperuricaemia and erythroplakia (159).
rs2078267	11	*SLC22A11 (Solute Carrier Family 22 Member 11)*	• The *SLC22A11* gene encodes the member 11 of organic anion transporter (OAT4) family (160). The encoded protein is an integral membrane protein involved in the sodium-independent transport and excretion of organic anions (161). • This transporter is primarily found in epithelial cells of the proximal tubule in the kidney and in the placenta, where it may act to excrete harmful organic anions (160). • GO annotations related to this gene include inorganic anion exchanger activity and sodium-independent organic anion transmembrane transporter activity (162). • Disease associated with the *SLC22A11* includes hyperuricaemia (162).

SNP	CHR	GENE	Gene function (from http://www.genecards.org/)
rs1165151	6	*SLC17A1 (Solute Carrier Family 17 Member 1)*	• The *SLC17A1* gene encodes the sodium-dependent phosphate transport protein 1 (NPT1) (163). NPT1 is important for the reabsorption of phosphate and the urate excretion in the kidney, which actively reabsorbs phosphate into cells via Na+ cotransport and excretes urate into the distal renal tubule (164). • Its related biological pathways involve the transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds, and uricosuric pathway (pharmacodynamics) (165). • GO annotation related to this gene includes symporter activity and phosphate ion transmembrane transporter activity. • Diseases associated with the *SLC17A1* gene include gout and hyperuricaemia (165).
rs1471633	1	*PDZK1 (PDZ Domain Containing 1)*	• The protein encoded by the *PDZK1* is the Na+/H+ exchange regulatory cofactor NHE-RF3, mediating the localisation of cell surface proteins (166). This protein is likely to influence urate transport indirectly by binding with the urate transporters URAT1, NPT1, and OAT4 (167). It also contributes to cholesterol metabolism by regulating the HDL receptor, namely scavenger receptor class B type 1 (168). • Its related biological pathways include the regulation of cystic fibrosis transmembrane conductance regulator (CFTR) activity and the uricosurics pathway (pharmacodynamics) (169). • GO annotation related to this gene includes PDZ domain binding and the transporter activity (169). • Diseases associated with the *PDZK1* gene include multiple myeloma (169).

SNP	CHR	GENE	Gene function (from http://www.genecards.org/)
rs1171614	10	*SLC16A9 (Solute Carrier Family 16 Member 9)*	• The protein encoded by *SLC16A9* gene is a proton-linked monocarboxylate transporter (MCT9) (170). The predicted function of SLC16A9 (MTC9) is a carnitine efflux transporter, which catalyses the transport of monocarboxylates across the plasma membrane (171). • Whether SLC16A9 (MTC9) directly contributes to the transport of urate is not very clear, but associations between this gene and another two metabolites were observed, namely DL-carnitine and propionyl-L-carnitine, which in turn were associated with serum urate, forming a triangle between SLC16A9, serum urate and related metabolites (151). • GO annotation related to this gene includes the symporter activity and the transmembrane transporter activity of monocarboxylic acid (172). • Diseases associated with this gene include gout (172).

1.3.2 Other genetic risk loci

For the remaining genetic risk loci identified from GWAS, their underlying biological functions in the regulation of serum urate have not been completely understood, but the functional annotation of the mapping genes highlights two broad pathways of glycolysis and inhibins/activins, which contribute to the biological processes of energy balance, cell proliferation, apoptosis, and immune response, as described below. Gene functions for other genetic risk loci of SUA levels are displayed in **Table 1-5** and **Figure 1-4**.

Table 1 - 5: Gene functions for other genetic risk loci identified in SUA GWAS.

SNP	Chr	Gene	Gene function (from http://www.genecards.org/)
rs1260326	2	*GCKR (Glucokinase Regulator)*	• The *GCKR* gene encodes a regulatory protein that inhibits glucokinase in liver and pancreatic islet cells by binding to the enzyme to form an inactive compound (173, 174). • Three biological pathways are related to this gene: the transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds, cyclin-dependent kinase (CDK) mediated phosphorylation and the removal of cell division cycle 6 (Cdc6) from the cellular nucleus (175). • GO annotation related to this gene includes enzyme binding and protein domain specific binding (175). • Diseases associated with the *GCKR* include fasting plasma glucose level and maturity-onset diabetes (175).
rs3741414	12	*INHBC (Inhibin Beta C Subunit)*	• The *INHBC* gene encodes a member of the transforming growth factor-beta (TGF-β) superfamily, that could be processed by proteolysis to produce a subunit of homodimeric and heterodimeric activin compound (176). The heterodimeric complex may inhibit the signalling of activin A (177). • The biological pathways related to this gene include the TGF-beta signalling pathway (KEGG) and signalling pathway regulating pluripotency of stem cells (178). • GO annotation of this gene includes cytokine activity and hormone activity (178).

SNP	Chr	Gene	Gene function (from http://www.genecards.org/)
rs675209	6	*RREB1 (Ras Responsive Element Binding Protein 1)*	• The product of *RREB1* gene is a zinc finger transcription factor binding to the RAS-responsive elements (RREs) of gene promoters (179). • This protein is involved in Ras/Raf-mediated cell differentiation by binding to the RRE of calcitonin gene promoter (180). • GO annotation related to this gene includes the binding of nucleic acid and RNA polymerase II core promoter sequence-specific DNA (181).
rs2941484	8	*HNF4G (Hepato-cyte Nuclear Factor 4 Gamma)*	• The *HNF4G* gene encodes the member 2 of nuclear receptor subfamily 2, also known as NR2A2. • The related biological pathways of this gene are the regulation of beta-cell development and gene expression (182, 183). • GO annotations related to this gene include the transcription factor activity, steroid hormone receptor activity and sequence-specific DNA binding (184). Diseases associated with this gene include maturity-onset diabetes (184).
rs6770152	3	*SFMBT1 (Scm-Like With Four Mbt Domains 1)*	• The protein encoded by the *SFMBT1* gene contains four malignant brain tumour repeat (mbt) domains (185). • It is likely involved in antigen recognition (186). • GO annotations related to this gene include transcription corepressor activity and histone binding (187).
rs642803	11	*OVOL1 (Ovo Like Transcripti onal Repressor 1)*	• The *OVOL1* gene encodes a putative zinc finger transcription factor which is very similar to homologous protein in mouse and drosophila (188). • Based on its known functions in these species, this protein is likely involved in hair formation and spermatogenesis (189). • GO annotation indicates this gene is likely involved in RNA polymerase II core promoter proximal region sequence-specific binding and transcriptional repressor activity (189).

SNP	Chr	Gene	Gene function (from http://www.genecards.org/)
rs6598541	15	*IGF1R (Insulin-Like Growth Factor 1 Receptor)*	• The product of this gene is an insulin-like growth factor I receptor (IGF1R), containing alpha and beta subunits (190). This receptor is overexpressed in most malignant tissues where it may act as an anti-apoptotic agent by enhancing cell survival (190). • The biological pathways related to this gene include the development IGF-1 receptor signalling and the mammalian target of rapamycin (mTOR) (191). • GO annotation for this gene includes protein kinase activity and protein binding (191).
rs729761	6	*VEGFA (Vascular Endothelial Growth Factor A)*	• The protein encoded by the *VEGFA* gene is the member A of the PDGF/VEGF growth factor family (192). • This growth factor is essential for both physiological and pathological angiogenesis and it induces the proliferation and migration of the vascular endothelial cell (192). This expression of this gene is upregulated in some known tumours and is also correlated with tumour stage and progression. Mutations of this gene have been associated with atherosclerosis and microvascular complications of diabetes 1 (MVCD1) and atherosclerosis (192). • GO annotation highlight the role of this gene in the protein homo-dimerisation activity (193).
rs10821905	10	*A1CF (APOBEC1 Compleme-ntation Factor)*	• The product of this gene contains three different RNA recognition motifs, belonging to the hnRNP R family of RNA-binding proteins (194). This complementation factor may be involved in the RNA editing or RNA processing events (194). • Its related pathways are the mRNA editing of C to U conversion and the gene expression (195). • GO annotations related to this gene include nucleotide/nucleic acid binding (195).

SNP	Chr	Gene	Gene function (from http://www.genecards.org/)
rs1178977	7	*BAZ1B (Bromod-main Adjacent To Zinc Finger Domain 1B)*	• This gene encodes an enzyme, tyrosine-protein kinase belonging to the bromodomain protein family (196). • The bromodomain is a structural characteristic of proteins involving in chromatin-dependent regulation of transcription (196). • GO annotation related to this gene includes protein tyrosine kinase activity and binding (197). • This gene is deleted in a developmental disorder, Williams-Beuren syndrome (197).
rs1394125	15	*UBE2Q2 (Ubiquitin Conjugatin-g Enzyme E2 Q2)*	• The protein encoded by this gene is one of the ubiquitin conjugating enzymes (also known as E2 enzymes) (198). • Its related pathways are class I major histocompatibility complex (MHC) mediated antigen processing and innate immune response (199). • GO annotations of this gene include ligase activity (199).
rs653178	12	*ATXN2 (Ataxin 2)*	• The *ATXN2* gene encodes the protein that is involved in the trafficking of epidermal growth factor receptor (EGFR) (200). It functions as a negative regulator of endocytic EGFR internalisation at the plasma membrane (200). • The involving biological pathways are the regulation of checkpoint kinases. • GO annotation related to this gene includes the poly (A) RNA binding and protein C-terminus binding (201). • Diseases associated with *ATXN2* include Spinocerebellar Ataxia 2 and late-onset Parkinson's disease (201).
rs7224610	17	*HLF (Hepatic Leukaemia Factor)*	• This gene encodes a member of the proline and acidic-rich (PAR) protein family, a subset of the bZIP transcription factors (202). • The encoded protein forms homodimers or heterodimers with other PAR family members and binds to the sequence-specific promoter elements to activate transcription (202). • GO annotations related to this gene include transcription factor activity, double-stranded DNA binding and sequence-specific DNA binding (203). • Leukaemia is associated with this gene (203).

SNP	Chr	Gene	Gene function (from http://www.genecards.org/)
rs2079742	17	*BCAS3 (Breast Carcinoma Amplified Sequence 3)*	• The product of the *BCAS3* gene is a microtubule associated cell migration factor (204). • This protein is involved in multiple biological processes, including angiogenesis, the regulation of cell polarity and directional endothelial cell migration, and the recruitment and activation of cell division control protein (Cdc42) (205). • GO annotation related to this gene includes chromatin binding and histone binding (205). • Disease associated with this gene is breast cancer (205).
rs2307394	2	*ORC4L (Origin Recogni-tion Complex Subunit 4)*	• This gene encodes the subunit 4 of the origin recognition complex (ORC) that is an essential complex for the initiation of DNA replication in the cell (206). • The biological pathways related to ORC4 are the regulation of DNA replication, the cyclin-dependent kinase (CDK) mediated phosphorylation and the removal of Cdc6 (207). • GO annotation related to this gene includes the nucleotide binding and the DNA replication origin binding (207). • Meier-Gorlin Syndromes are associated with this gene (207).
rs17050272	2	*INHBB (Inhibin Beta B Subunit)*	• The *INHBB* gene encodes a subunit of the transforming growth factor-beta (TGF-β) superfamily that could be processed by proteolysis to generate a subunit of the dimeric activin and inhibin protein complexes (208). • These complexes respectively activate and inhibit the follicle stimulating hormone secretion activity of the pituitary gland (208). This gene is involved in the TGF-β signalling pathway (KEGG) and signalling pathways regulating pluripotency of stem cells (209). • GO annotation related to this gene includes the growth factor activity and protein homodimerisation activity (209). • Diseases associated with this genes are eclampsia and pre-eclampsia (209).

SNP	Chr	Gene	Gene function (from http://www.genecards.org/)
rs7193778	16	*NFAT5 (Nuclear Factor Of Activated T-Cells 5)*	• The *NFAT5* gene encodes a transcription factor involved in the transcription of inflammatory gene and the regulation of activated T cells family (210). • GO annotation of this gene includes sequence-specific DNA binding, transcriptional activator activity, and RNA polymerase II core promoter proximal region sequence-specific binding (211). • Diseases associated with this gene are kidney papillary necrosis and spinocerebellar ataxia (211).
rs10480300	7	*PRKAG2 (Protein Kinase AMP-Activated Non-Catalytic Subunit Gamma 2)*	• The *PRKAG2* gene encodes the non-catalytic regulatory gamma subunit of the AMP-activated protein kinase (AMPK) (212). • AMPK is an important energy-sensing enzyme that monitors cellular energy status by inactivating key enzymes involved in regulating *de novo* biosynthesis of fatty acid and cholesterol (212). • GO annotation of this gene includes protein kinase activator activity and protein kinase binding (213). • Mutations in this gene have been associated with Wolff-Parkinson-White syndrome, familial hypertrophic cardiomyopathy, and glycogen storage disease of the heart (213).
rs7188445	16	*MAF (MAF BZIP Transcription Factor)*	• The protein encoded by this gene is a DNA-binding, leucine zipper-containing transcription factor, acting as a homodimer or as a heterodimer (214). • This transcriptional factor involves in the regulation of several cellular processes, including embryonic lens fibre cell development, chondrocyte terminal differentiation, and increased T-cell susceptibility to apoptosis (214). • GO annotation related to this gene includes sequence-specific DNA binding, transcriptional activator activity, and RNA polymerase II transcription regulatory region sequence-specific binding (215). • Diseases associated with MAF are Ayme-Gripp Syndrome and multiple-type cataract (215).

SNP	Chr	Gene	Gene function (from http://www.genecards.org/)
rs17786744	8	*STC1 (Stanniocalcin 1)*	• The *STC1* gene encodes a secreted, homodimeric glycoprotein that may have autocrine or paracrine functions (216). • The protein is likely involved in the regulation of renal and intestinal calcium and phosphate transport, cell metabolism, or cellular calcium/phosphate homeostasis (216). • GO annotation for this gene includes the activity of hormone (217). • Diverticulitis of colon is associated with the mutations of this gene (217).
rs164009	17	*PRPSAP1 (Phosphoribosyl Pyrophosphate Synthetase Associated Protein 1)*	• The protein coded by *PRPSAP1* gene seems to play a negative regulatory role in 5-phosphoribose 1-diphosphate synthesis (218). GO annotations related to this gene include magnesium ion binding and enzyme inhibitor activity (219).
rs11264341	1	*TRIM46 (Tripartite Motif Containing 46)*	• The product of this gene is a tripartite motif-containing protein, which is involved in the biological processes of interferon gamma signalling and innate immune system (220, 221)
rs17632159	5	*TMEM171 (Transmembrane Protein 171)*	• *TMEM171* is a transmembrane protein coding gene but its function is not clear (222, 223).

2 AIMS AND OBJECTIVES

2.1 Introduction

The role of uric acid has been explored in a large number of observational studies in relation to a multitude of health outcomes. Apart from gout and urolithiasis (224, 225), compelling evidence supports the associations between high SUA level and the increased risk of non-crystal deposition disorders, including metabolic syndrome, hypertension, chronic renal diseases, and cardiovascular diseases (CVDs) (225-227). Although there are considerable research efforts into understanding the pathogenic role of uric acid in these non-crystal deposition disorders, their causal relationships have not been established. It was argued that either these associations are confounded by other risk factors, such as obesity and hypertension, or they represent reverse causality. Genetic determinants play a substantial role in the regulation of SUA levels and genetic studies among twins and families reported a substantial heritable component of SUA level with an estimated heritability of 40-70% (228, 229). The genetic component of SUA level has been explored in several GWAS (151, 230, 231) and the wealth of GWAS findings allows the incorporation of genetic variant(s) as an instrument (s) which can be used to separate causal associations from non-causal ones, given that the genotype is generally independent of environmental exposures and the transmission of genetic information is usually unidirectional. Therefore, investigating the associations between SUA genetic risk loci and disease outcomes might provide evidence for the hypotheses which link uric acid to clinical disorders.

2.2 Aim and objectives

The overarching aim of this project is to employ the Phenome-wide association study (PheWAS) approach along with other complementary methodologies to investigate the role of uric acid in a wide range of disease outcomes by using data from the UK Biobank.

The specific objectives are:

(i) To conduct an umbrella review of meta-analyses of observational studies, meta-analyses of randomised controlled trials (RCTs) and Mendelian randomisation (MR) studies on the associations between SUA level and multiple health outcomes. This umbrella review was performed to summarise the range of related health outcomes, present the magnitude, direction and significance of the identified associations and

effects, assess the potential biases and identify which of the observed associations are robust. The findings of umbrella review are presented in ***Chapter 3***.

(ii) To perform a MR-PheWAS analysis by using the interim release data of UK Biobank to provide an overview of the disease outcomes associated with the SUA genetic risk loci. The SUA genetic risk loci were employed as individual instruments. The phenome framework was defined by the PheCODE schema. PheWAS was performed first to identify any association across the SUA genetic risk loci and the phenome; the Mendelian randomisation (MR) design and the HEIDI (heterogeneity in dependent instruments) test were then applied to distinguish the PheWAS associations that were due to causality, pleiotropy or genetic linkage. The findings of MR-PheWAS are presented in ***Chapter 5***.

(iii) To validate the MR-PheWAS findings, an advanced Phenome-wide Mendelian randomisation (PWMR) analysis is performed by using data from the full UK Biobank cohort. A weighted polygenic risk score (GRS), incorporating effect estimates of multiple genetic risk loci, was employed as a proxy of the SUA level. The phenome framework was also defined by an alternative tree-structured phenotypic model. Any novel findings and/or replication of the MR-PheWAS findings were further described and explored in the PWMR analysis. The findings of PWMR are presented in Chapter 6.

3 UMBRELLA REVIEW ON SUA LEVEL

3.1 Summary

The aim of this chapter is to assess if the available evidence is robust enough to indicate any causal effect of SUA level on the related health outcomes. I performed an umbrella review to collect and evaluate the evidence from systematic reviews and meta-analyses of observational studies, meta-analyses of RCTs and Mendelian randomisation studies systematically. I identified 136 health outcomes that were examined in relation to SUA levels across three study types, including anthropometric outcomes, cardiovascular diseases, metabolic diseases, kidney disorders, various cancers, and neurocognitive diseases. I adopted a set of criteria to assess the credibility of the observed associations. After assessment, no association from observational studies was classified as convincing, and associations with five health outcomes (heart failure, hypertension, impaired fasting glucose or diabetes, chronic kidney disease, coronary heart disease mortality) were classified as highly suggestive; only one outcome (nephrolithiasis) from meta-analyses of RCTs and one outcome (gout) from Mendelian randomisation studies presented convincing evidence, indicating a causal effect of high SUA level on increased disease risk.

Despite a few hundred systematic reviews, meta-analyses, and Mendelian randomisation studies exploring 136 unique health outcomes, convincing evidence of a clear role of SUA level only exists for gout and nephrolithiasis. Umbrella reviews focus on the evidence from existing meta-analyses and therefore outcomes that have not been assessed in a meta-analysis are not included in this review, which represents a weakness of this study. The available evidence does not support any change in existing clinical recommendations in relation to the clinical management of SUA levels.

This chapter has been published in the *British Medical Journal* (BMJ) cited as "*Li, X., Meng, X., Timofeeva, M., Tzoulaki, I., Tsilidis, K.K., Ioannidis, P.A., Campbell, H. and Theodoratou, E., 2017. Serum uric acid levels and multiple health outcomes: umbrella review of evidence from observational studies, randomised controlled trials, and Mendelian randomisation studies. BMJ, 2357-2376."* (232).

For this published work, I conducted the systematic literature review and data extraction as the main investigator. The extracted data was checked by the second investigator (*Meng, X.*). I performed the statistical analysis, wrote the manuscript and revised the paper according to the comments given by the peer reviewers and the BMJ editorial committee. *Theodoratou, E.*

and *Campbell, H.* conce*iv*ed the study and *Ioannidis, P.A.* contributed to the design. *Campbell, H., Ioannidis, P.A., Tsilidis, K.K., Timofeeva, M.,* and *Tzoulaki, I.* critically reviewed the manuscript and contributed important intellectual content.

Serum uric acid and multiple health outcomes: an umbrella review of evidence from observational studies, clinical trials and Mendelian randomisation studies

AUTHORS Xue Li[1], Xiangrui Meng[1], Maria Timofeeva[2], Ioanna Tzoulaki[3], Konstantinos K. Tsilidis[3,4], John P.A. Ioannidis[5], Harry Campbell[1], Evropi Theodoratou[1,6]*

1 Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom

2 Colon Cancer Genetics Group and Academic Coloproctology, Institute of Genetics and Molecular Medicine, University of Edinburgh and MRC Human Genetics Unit Western General Hospital Edinburgh, Edinburgh, United Kingdom

3 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

4 Department of Hygiene and Epidemiology, University of Ioannia School of Medicine, Ioannina, Greece

5 Department of Medicine, Stanford Prevention Research Center, Stanford School of Medicine, Stanford, CA, 94305, USA; Department of Health Research and Policy, Stanford School of Medicine, Stanford, CA, 94305, USA; Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, 94305, USA

6 Edinburgh Cancer Research Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom

***Corresponding author**

Evropi Theodoratou, Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom, EH8 9AG; E-mail: E.Theodoratou@ed.ac.uk, Tel: (+44) 131-650-6194

ABSTRACT

Objective To map the diverse health outcomes associated with serum uric acid (SUA) and assess the credibility of evidence in favour of causal association.

Design Umbrella review of the evidence from systematic reviews and meta-analyses of observational studies, meta-analyses of randomised controlled trials (RCTs) and Mendelian randomisation (MR) studies.

Data sources Medline, EMBASE, Cochrane database of systematic reviews and screening of citations and references.

Eligibility criteria Systematic reviews and meta-analyses of observational studies that examined associations between SUA and health outcomes; meta-analyses of RCTs that investigated health outcomes related to SUA-lowering therapy; MR studies that explored the causal associations of SUA with health outcomes.

Results 47 articles reporting 144 meta-analyses of observational studies (76 unique outcomes), 8 articles reporting 31 meta-analyses of RCTs (20 unique outcomes) and 36 articles reporting 107 MR studies (56 unique outcomes) met the eligibility criteria. Across all three study types, 136 unique health outcomes (including anthropometric outcomes, cardiovascular diseases, metabolic diseases, kidney disorders, various cancers, and neurocognitive diseases) were reported. 16 of 76 unique outcomes in meta-analyses of observational studies had $p<10^{-6}$, 8 of 20 unique outcomes in meta-analyses of RCTs had $p<10^{-3}$ and 4 of 56 unique MR outcomes had $p<0.01$. Large between-study heterogeneity was very common (80.3% and 45.0% in meta-analyses of observational studies and of RCTs respectively). 42 of 76 (55.3%) meta-analyses of observational studies and 7 of 20 (35.0%) meta-analyses of RCTs showed evidence of small study effects or excess significance bias. No SUA-health outcome associations from meta-analyses of observational studies were classified as convincing; five associations were classified as highly suggestive (increased risk of heart failure, hypertension, impaired fasting glucose or diabetes, chronic kidney disease (CKD), coronary heart disease mortality with high SUA levels). Only one outcome from RCTs (decreased risk of nephrolithiasis recurrence with SUA-lowering treatment) had $p<10^{-3}$, 95% prediction interval excluding the null and no large heterogeneity or bias. Only one outcome from MR studies (increased risk of gout with high SUA levels) presented convincing evidence. Hypertension and CKD showed concordant evidence in meta-analyses of observational studies and in some (but not all) meta-analyses of RCTs with respective intermediate or surrogate outcomes, but they were not statistically significant in MR studies.

Conclusions Despite a few hundred systematic reviews, meta-analyses and MR studies exploring 136 unique health outcomes, convincing evidence of a clear role of SUA exists for only gout and nephrolithiasis.

What is known or unknown on this topic

- Observational studies have suggested that high SUA levels are associated with a wide range of outcomes including cardiovascular and metabolic diseases (increased risk) or neurological diseases (decreased risk).
- However, it remains to be determined whether these observed associations are causal.
- Clinical trials of SUA lowering have shown that xanthine oxidase inhibition decreases blood pressure and improves renal function.
- There is still much debate as to whether SUA is simply a marker of xanthine oxidase activity or a causal factor involved in systemic inflammation.

Added value by this study

- We present here a comprehensive overview and assessment of the existing evidence from multiple sources (including meta-analyses of observational studies, meta-analyses of RCTs and MR studies) for a wide range of health outcomes related to SUA.
- We identified 136 health outcomes that were examined in relation to SUA but based on our evidence assessment convincing evidence of a clear role of SUA exists only for gout and nephrolithiasis.
- Associations between SUA and five additional health outcomes (heart failure, hypertension, impaired fasting glucose or diabetes, chronic kidney disease, coronary heart disease mortality) were classified as highly suggestive.
- There is a notable gap between observational studies and RCTs and MR studies.

Implications of all available evidence

- This study raises large uncertainty about the potential therapeutic benefits of SUA-lowering therapy beyond gout.
- There is no adequate evidence against lowering SUA in gout patients in relation to an increased risk of neurological diseases.
- The causal effect of SUA on the health outcomes with highly suggestive evidence might be worthy of further investigation.

3.2 Introduction

Uric acid was thought to be a biologically inert waste product from purine metabolism, until Garrod *et al* discovered in the early 1800s that elevated SUA level was the cause of gout (233). Subsequently, associations of uric acid with cardiovascular and renal disorders were also observed (234). These associations were explored in a number of prospective studies, but yielded conflicting results and, therefore, the causal role of uric acid in these diseases was widely questioned (235-238). It was argued that either these associations are confounded by other risk factors, such as obesity and hypertension, or they represent reverse causality (236, 239). These inconclusive findings led to a shift of interest away from uric acid and asymptomatic hyperuricaemia was not considered as indication for SUA-lowering therapy in patients with cardiovascular and renal diseases (56, 124).

New findings have fuelled enthusiasm to address this long-standing controversy (240). Recent epidemiological studies have explored associations of uric acid with a wide range of disease conditions, including cardiovascular diseases (CVD), metabolic syndrome, diabetes, cancer and some intermediate phenotypes or biomarkers (227). In an attempt to understand the possible underlying mechanisms, laboratory studies have found that uric acid is potentially involved in multiple biological processes including oxidation stress, systemic inflammation and intra-hepatic fructose metabolism, which are mechanisms that could be linked to the development of CVD and metabolic syndrome (241-243). Alternatively, uric acid may only present a marker of high oxidative stress associated with increased xanthine oxidase activity, instead of an active agent in the pathogenic processes (244). Finally, taking into account the antioxidant properties of uric acid (acting as a free radical scavenger), its potential mechanistic roles on these disorders may be complex (245).

In view of the potential importance of uric acid, assessing the credibility of the observed evidence may have implications both for clinical practice and public health. It is recognised that different types of studies have specific strengths and weaknesses that can be seen as complementary (see Box 1). An umbrella review, which collects and evaluates evidence from multiple resources systematically, might therefore help clarify the composite literature. Here, we have performed an umbrella review of meta-analyses of observational studies, meta-analyses of randomised controlled trials (RCTs) and Mendelian randomisation (MR) studies on associations between SUA and multiple health outcomes. As shown in Box 1, these three lines of evidence can be complementary. The overarching aim of this study is to provide an overview of the spectrum of diseases related to SUA and to assess the evidence from multiple sources systematically. In particular, we summarised the range of related

health outcomes, presented the magnitude, direction and significance of the reported associations and effects, assessed the potential biases and identified which associations and effects have the most convincing evidence.

Box 1: observational studies, randomised clinical trials, and Mendelian randomisation studies

- An observational study aims to examine the association between an exposure and an outcome, and tests whether the association is caused by chance, bias or confounding. However, conventional observational studies typically may suffer from problems of residual confounding, undetected bias or reverse causality which may generate associations that are not reliable indicators of causality.

- A randomised clinical trial (RCT) is an approach to obtain evidence of a causal effect of a treatment or intervention on a disease process, and it eliminates many of the biases and confounding factors that are present in observational studies. However, this study design is also subject to limitations, including non-adherence to the assigned intervention, limited external validity, short-term intervention effects, and non-retention, which can all render the results invalid or questionable. In addition, high costs and ethical concerns can also limit the application of RCTs in scientific research.

- A Mendelian randomisation (MR) study provides a cost-effective analogy to an RCT by using genetic variants as proxies to test the causality of an association between the exposure and outcome. MR is not influenced by the confounding inherent in observational studies and not seriously affected by reverse causality, but it does rely on several assumptions (the genetic instruments should be associated with the exposure of interest, they should not be associated with known confounders, and they should affect the outcome solely through the exposure) that can be hard to identify and control. This approach may also lack power especially when the proportion of trait variance explained by the genetic instruments is small.

- In summary, although none of these study types are infallible, all of them are able to provide useful information in relation to causal inference and they can complement each other in order to achieve increasing certainty about causality.

3.3 Methods

3.3.1 Literature search and selection criteria

We systematically searched Medline, EMBASE and the Cochrane database from inception to July 17, 2016 using a comprehensive search strategy (**Supplementary Table 3-1**) to identify systematic reviews and meta-analyses of observational studies, meta-analyses of RCTs, and MR studies. All identified publications went through a three-step parallel review of title, abstract and full text (performed by XL and XM) based on pre-defined inclusion and exclusion criteria.

We included: (i) systematic reviews and meta-analyses of observational studies that examined associations between SUA levels (or hyperuricaemia) and health outcomes; (ii) meta-analyses of RCTs that investigated health outcomes related to SUA-lowering therapy (intervention with one or a combination of two or more SUA-lowering drugs versus placebo or no treatment) including xanthine oxidase inhibitors (allopurinol, febuxostat, or oxypurinol), uricosuric agents (probenecid, benzbromarone, thiazides, or citrates), and uricase analogues (pegloticase or rasburicase); and (iii) individual MR studies that explored SUA (or hyperuricaemia) associations in relation to health outcomes by using genetic instruments influencing SUA levels. The identified health outcomes included a wide range of diseases, intermediate phenotypes and biomarkers. Studies investigating associations between gout and health outcomes were not included. Meta-analyses of RCTs that used non-drug interventions, such as dietary or lifestyle interventions were excluded. We further excluded animal and laboratory studies, meta-analyses on the prevalence of gout and hyperuricaemia and meta-analyses of RCTs that focused on pharmacological parameters, safety and effects of reducing SUA levels without investigating other health effects.

3.3.2 Data extraction

Data were extracted by one investigator (XL) and then checked by a second investigator (XM). For each eligible study, we abstracted the PubMed ID, first author name, journal, year of publication, study population, number of studies included, and outcomes investigated. For meta-analyses investigating more than one health outcome, each outcome was recorded separately. For meta-analyses of observational studies and of RCTs, we extracted the reported summary risk estimates (risk ratio, odds ratio, hazard ratio or mean difference) with the 95% confidence intervals (CIs) and the corresponding number of case and control participants. Furthermore, for each unique outcome, data from the individual component

studies that were included in the meta-analyses were also extracted for further analysis. This second level abstraction included data on study design, number of cases, total number of study participants, the relative risk estimates and 95% CIs for each component study. When more than one meta-analysis existed for the same outcome in the same population, individual component data were extracted from the most recent and largest meta-analysis. In a few exceptions where the most recent was not also the largest meta-analysis, we explored the reason for this discrepancy: if the most recent included prospective studies and the largest one had fewer prospective studies plus some retrospective data, we kept the one with the largest amount of prospective data; otherwise we kept the largest meta-analysis. For MR studies, we abstracted data on study population, sample size, genetic instruments, the variance of SUA explained by the genetic instruments (R^2) and MR effect estimates (odds ratio, hazard ratio, mean difference or regression coefficient β), standard deviation (SD) of SUA levels and SD of continuous outcomes.

3.3.3 Data analysis

For systematic reviews, we performed descriptive analyses and presented the authors' conclusions. For each unique meta-analysis of observational studies and of RCTs, we estimated a number of metrics including (i) the summary effect and 95% CI using a random-effect model (DerSimonian & Laird method) (246); (ii) the heterogeneity among studies (Q statistic and I^2 metric with 95% CI); (iii) the 95% prediction interval (PI) to predict the range of effect size that would be expected in a new original study, after accounting for both the heterogeneity among individual studies and the uncertainty of the summary effect estimated in the random-effect model; the calculation of 95% PI is based on the predicted distribution derived from a function of the degree of heterogeneity, the number of studies included and the within-study standard errors (247, 248); (iv) the presence of small study effects by using the Egger's regression asymmetry test to investigate if small studies tend to give larger estimates of effect size than large studies (significance threshold $p<0.10$) (249); (v) the excess significance test to assess if the observed number (O) of studies with significant results was greater than the expected number (E) using the chi-square test: $A = [(O-E)^2/E + (O-E)^2/(n-E)] \sim \chi^2$ (significance threshold $p<0.10$) (250, 251). For the excess significance test the expected number (E) of studies with significant findings was calculated by using the sum of statistical power estimated for each component study. The statistical power of each component study was calculated with an algorithm that uses a non-central *t* distribution, by assuming the true effect size to be the same as that of the largest component study (with smallest variance) in the meta-analysis (252). If the type of metric in a meta-analysis was

mean difference, we firstly calculated *Cohen's d* by weighing the pooled SD based on the sample size of individual studies. We then transformed *Cohen's d*, *Hedges g* and other standardised mean difference metrics to odds ratios (253). We compared the results reported in overlapping meta-analyses to evaluate their concordance in terms of the direction and statistical significance of the observed associations. All statistical analyses were conducted in Stata version 14.0.

For MR studies, we didn't conduct quantitative syntheses due to the extensive differences of genetic instruments used in the identified studies. Instead, we performed and present here a descriptive analysis of the individual studies. When more than one MR study was conducted for the same outcome, we compared the concordance of the findings in relation to the direction and statistical significance of the reported association and retained the MR study with largest number of cases and participants for further analysis and comparison. We performed a power calculation for the largest MR studies by using the non-centrality parameter (NCP) based approach, if all necessary information required for calculation was provided (i.e., sample size, number of cases, R^2, estimates of association, SD of continuous outcomes and SD of SUA levels) (254). For MR study with missing R^2, we used the R^2 from other MR studies that used the same genetic variants as instruments to perform a crude power estimation.

3.3.4 Credibility assessment

Evidence from meta-analyses of observational studies with nominally statistically significant summary results ($p<0.05$) was classified into four categories (class I, II, III, IV) as previously proposed (255): (i) convincing (class I) evidence was assigned to associations, which had a statistical significance of $p<10^{-6}$, included more than 1,000 cases (or more than 20,000 participants for continuous outcomes), had the largest component study reporting a statistically significant result ($p<0.05$), had a 95% PI that excluded the null, did not have large heterogeneity ($I^2<50\%$), and showed no evidence of small study effects ($p>0.10$) and of excess significance bias ($p>0.10$); (ii) highly suggestive (class II) evidence was assigned to associations, which reported a statistical significance of $p<10^{-6}$, included more than 1,000 cases (or more than 20,000 participants for continuous outcomes), and had the largest component study reporting a statistically significant result ($p<0.05$); (iii) suggestive (class III) evidence was assigned to associations, which reported a statistical significance of $p<10^{-3}$ with more than 1,000 cases (or more than 20,000 participants for continuous outcomes); (iv) weak (class IV) evidence was assigned to the remaining statistically significant associations with $p<0.05$. For each association in the convincing or highly suggestive categories we re-assessed the evidence after excluding the retrospective and case-control studies (in an

attempt to address reverse causality). Finally, for each association in the convincing category we re-assessed the evidence after we examined each meta-analysis in depth ourselves by assessing the eligibility of the included studies as well as verifying the data used in the meta-analysis.

Evidence from meta-analyses of RCTs was assessed in terms of the statistical significance of the summary effect ($p<10^{-3}$, $10^{-3}\leq p<0.05$, $p\geq 0.05$), 95% PI (excluding the null or not), and presence of large heterogeneity ($I^2>50\%$), small study effects ($p>0.10$) and excess significance ($p>0.10$). We also noted the conclusions from any evidence classification (GRADE (256) or equivalent system) applied by the original meta-analyses. Finally, evidence from individual MR studies was assessed in terms of statistical significance of the MR effect estimate ($p<0.01$) and of the statistical power (>80%) (257).

For overlapping outcomes that were investigated in meta-analyses of observational studies and/or meta-analyses of RCTs and/or individual MR studies, we examined if the direction and statistical significance of the associations were reported concordantly across the different study types. We noted the overlapping outcomes that were graded as class I-II in meta-analyses of observational studies and had 95% PI excluding the null in meta-analyses of RCTs. For these outcomes we also presented the evidence from MR studies if available.

3.3.5 Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in the design and implementation of the study. No patients were asked to advice on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

3.4 Results

3.4.1 Literature review

Overall, the parallel reviews identified 4,608 publications across three databases. After application of the inclusion/exclusion criteria, 101 publications were selected for inclusion (**Figure 3-1**). Specifically, 15 systematic reviews of observational studies were reported in 10 articles (**Supplementary Table 3-2**) (258-267); 144 meta-analyses of observational studies were reported in 47 articles (**Supplementary Table 3-3**) (16, 268-313); 31 meta-analyses of RCTs were reported in 8 articles (**Supplementary Table 3-4**) (50, 314-320); and 107 MR studies were reported in 36 articles (**Supplementary Table 3-5**) (321-356). Across all three study types, 136 unique outcomes were reported.

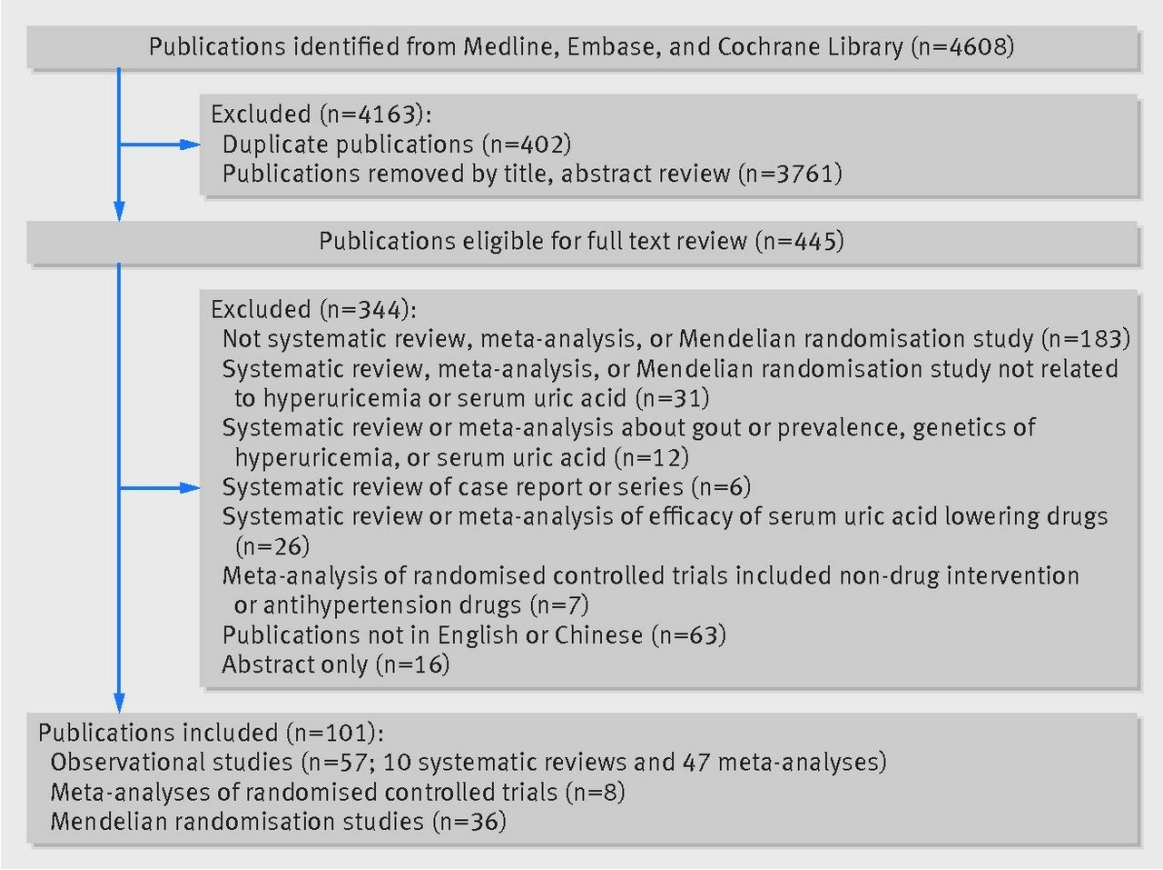


Figure 3 - 1: Study flowchart.

3.4.2 Meta-analyses of observational studies

One hundred and forty-four meta-analyses of observational studies were identified in total (**Supplementary Table 3-3**). The median number of studies included in meta-analyses was 5

(range: 2-31) and the median number of participants and median number of cases were 7,932 (range: 129-1,017,810) and 1,176 (range: 49-34,370) respectively. More than one meta-analysis was conducted for 16 outcomes (**Supplementary Table 3-3**). The direction and statistical significance of the reported associations in overlapping meta-analyses were concordant for 14 (87.5%) outcomes: atrial fibrillation incidence (n=3) (268, 280, 310), coronary heart disease (CHD) (n=4) (16, 300, 304, 311), hypertension incidence (n=3) (272, 302, 313), stroke incidence (n=2) (276, 303), diabetes (n=3) (277, 278, 307), chronic kidney disease (CKD) (n=3) (282, 283, 305), mild cognitive impairment (MCI) (n=2) (286, 308), Parkinson's disease (PD) (n=3) (286, 287, 309), multiple sclerosis (n=2) (288, 306), CHD mortality (n=3) (16, 300, 304), CVD mortality (n=2) (293, 312), stroke mortality (n=2) (276, 303), all-cause mortality in heart failure patients (n=2) (271, 295), and all-cause mortality in the general population (n=2) (293, 312). Discordance in the statistical significance was present for 2 outcomes: diabetic neuropathy (n=2) (279, 281) and Alzheimer's disease (AD) (n=4) (285, 286, 301, 308).

After removing the overlapping meta-analyses (which were conducted in the same population for the same outcome), 76 unique meta-analyses were retained reporting a wide range of outcomes (**Table 3-1**): cardiovascular outcomes (n=13), diabetes related outcomes (n=9), kidney disorders (n=7), neurocognitive disorders (n=11), cancer outcomes (n=6), all-cause or cause-specific mortality (n=22), and other outcomes (n=8). Overall, 58 (76.3%) of the 76 non-overlapping meta-analyses reported nominally statistically significant summary results ($p<0.05$). In **Supplementary Figure 3-1** and **3-2**, we plot the summary effects of the unique meta-analyses of observational studies. Of these, 12 (92.3%) meta-analyses in cardiovascular outcomes, 8 (88.9%) meta-analyses in diabetes related outcomes, all 7 (100.0%) meta-analyses in kidney disorders, one (9.1%) meta-analyses in neurocognitive disorders, one (16.7%) meta-analyses in cancer outcomes, 15 (68.2%) meta-analyses in all-cause and cause-specific mortality, and 6 (75.0%) meta-analyses in other outcomes reported summary estimates with $p<0.05$ and suggested that high levels of SUA were associated with an increased risk of disease. In addition, 7 (63.6%) meta-analyses in neurocognitive disorders, and 1 (12.5%) meta-analysis in other outcomes (composite of adverse outcomes (death or major cardiovascular event [MACE]) in acute ischaemic stroke patients) reported summary estimates with $p<0.05$ and suggested inverse associations with SUA.

We then applied our evidence classification criteria. Sixteen (21.1%) meta-analyses had $p<10^{-6}$; 10 (13.2%) meta-analyses had a 95% PI that excluded the null; 27 (35.5%) meta-analyses had more than 1,000 cases (or more than 20,000 participants for continuous

outcomes); 15 (19.7%) meta-analyses had no large heterogeneity (I^2<50%); 34 (44.7%) meta-analyses had neither small study effects nor excess significant bias. Based on these metrics, only one of 76 (1.3%) outcomes presented convincing evidence (class I: stroke mortality in general population); 7 (9.2%) outcomes presented highly suggestive evidence (class II: heart failure incidence, hypertension incidence, impaired fasting glucose or diabetes, CKD incidence, CHD mortality, all-cause mortality in heart failure patients, and non-alcoholic fatty liver disease); 9 (11.8%) outcomes presented suggestive evidence (class III: atrial fibrillation, CHD incidence, CVD, prehypertension, medium term MACE, T2DM, CVD mortality, CKD mortality, death or cardiac events). The remaining 41 (53.9%) nominally significant outcomes presented weak evidence (class IV).

We performed in-depth examination and re-assessed the meta-analyses of stroke mortality (276) (class I) and found that data from the largest study was incorrect (the events represented stroke incidence cases rather than stroke deaths and the included study had not published data on stroke mortality) (357). Furthermore, the data from two individual studies reported comparisons of SUA categories that differed from other studies (the highest sextile versus the second or third sextile rather than the lowest) (358, 359), and a fourth study had been included using only data on ischaemic stroke deaths but missing the data on haemorrhagic stroke deaths (360). When we excluded the stroke incidence study, used the proper comparison for the other two studies, and added the missing data in the fourth study, the association with stroke mortality was found to be non-significant (**Table 3-2**). For the highly suggestive outcomes (class II), when we limited the data to prospective cohort studies, all associations retained their ranking, except for all-cause mortality in heart failure patients and non-alcoholic fatty liver disease that were downgraded to class III (**Table 3-2**).

Table 3 - 1: Health outcomes and evidence class reported in meta-analyses of observational studies.

Outcome	Population	Study design included in MA	Comparison	N studies	N participants	N cases	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval	Evidence class[§]
Cardiovascular outcomes														
Atrial fibrillation (AF) (268)	General	Prospective cohort	Hyper vs normal	6	426,159	7,595	RR	1.49 (1.24, 1.79)	2.50E-05	79 (42, 89)	0.01	0.22	0.87-2.53	III
AF recurrence (269)	AF patients	Prospective/ retrospective cohort	Hyper vs normal	4	1,298	393	OR	1.52 (1.19, 1.94)	8.25E-04	89 (61, 95)	0.72	0.26	0.27-7.01	IV
CHD incidence (16)	General	Prospective cohort	Hyper vs normal	13	70,382	6,666	aRR	1.13 (1.05, 1.21)	7.70E-04	38 (0, 64)	0.27	<0.001	0.94-1.34	III
CVD (270)	Hypertensive patients	Prospective cohort	Hyper vs normal	6	19,546	1,054	aHR	1.17 (1.07, 1.27)	3.56E-04	67 (0, 84)	0.05	0.04	0.90-1.52	III
Heart failure incidence (271)	General	Prospective cohort	Hyper vs normal	5	427,917	10,171	HR	1.65 (1.41, 1.94)	1.77E-09	72 (7, 86)	0.49	0.31	1.05-2.61	II
Hypertension incidence (272)	General	Prospective cohort /nested case-control	Hyper vs normal	17	71,630	18,751	aRR	1.48 (1.33, 1.65)	3.99E-12	79 (65, 85)	0.06	NP	0.99-2.23	II
Prehypertension (273)	General	Cross-sectional	Highest vs lowest SUA category	8	44,095	20,832	OR	1.84 (1.42, 2.38)	4.88E-06	91 (86, 94)	0.10	NP	0.81-4.01	III
Left atrial thrombus or spontaneous echo contrast (LATH/LASEC) (268)	Patients with heart diseases	Prospective/ retrospective cohort	Highest vs lowest SUA category	6	2,381	241	OR	1.59 (1.13, 2.23)	7.51E-03	85 (66, 91)	0.02	NP	0.54-4.70	IV
Major adverse cardiovascular events (MACE)† (274)	Patients after PCI	Prospective/ retrospective cohort	Hyper vs normal	2	3,054	NA	RR	1.78 (1.26, 2.52)	1.16E-03	NA	NA	NP	NA	IV

Outcome	Population	Study design included in MA	Comparison	N studies	N participants	N cases	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval	Evidence class§
Medium-term MACE (275)	Acute myocardial infarction (AMI) patients	Prospective/ retrospective cohort	Highest vs lowest SUA category	4	4,299	1,240	OR	1.93 (1.36, 2.74)	2.56E-04	74 (0, 89)	0.81	NP	0.46-8.21	III
Short-term MACE (275)	AMI patients	Prospective/ retrospective cohort	Highest vs lowest SUA category	7	6,470	787	OR	2.46 (1.84, 3.27)	1.93E-09	63 (0, 82)	0.25	NP	1.06-5.71	IV
Stroke (270)	Hypertensive patients	Prospective/ retrospective cohort	Continuous SUA level	3	9,978	217	aHR	1.11 (0.98, 1.26)	0.10	70 (0, 89)	0.22	0.06	0.26-4.77	NS
Stroke incidence (276)	General	Prospective cohort	Highest vs lowest SUA category	5	24,548	1,290	aRR	1.22 (1.02, 1.46)	0.03	53 (0, 75)	0.03	NP	0.73-2.04	IV
Diabetes related outcomes														
Type 2 diabetes (T2DM) (277)	General	Prospective/ retrospective cohort	1 mg/dL SUA increase	11	42,834	3,305	RR	1.17 (1.09, 1.25)	8.97E-06	75 (54, 84)	0.07	0.002	0.92-1.47	III
Impaired fasting glucose or T2DM (278)	General	Prospective/ retrospective cohort	Highest vs lowest SUA category	12	62,834	6,340	RR	1.57 (1.39, 1.77)	1.12E-12	42 (0, 67)	0.09	NP	1.10-2.23	II
Diabetes incidence† (270)	Hypertensive patients	Prospective/ retrospective cohort	Hyper vs normal	2	8,247	564	aHR	1.84 (1.02, 3.30)	0.04	NA	NA	0.42	NA	IV
Diabetic nephropathy (279)	T2DM patients	Case-control	Categorical SUA level	3	3,166	196	OR	1.72 (1.07, 2.76)	0.03	84 (12, 93)	0.04	NP	0.01-382.85	IV
Diabetic microvascular complications (279)	T2DM patients	Case-control	Categorical SUA level	5	4,513	854	OR	1.42 (1.11, 1.83)	6.00E-03	83 (61, 90)	0.08	NP	0.68-2.95	IV

Outcome	Population	Study design included in MA	Comparison	N studies	N participants	N cases	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval	Evidence class[§]
Diabetic vascular complications (280)	T2DM patients	Case-control	Categorical SUA level	6	5,017	967	OR	1.27 (1.11, 1.45)	4.86E-04	77 (57, 86)	0.02	0.51	0.87-1.86	IV
Diabetic peripheral neuropathy (281)	Diabetic patients	Cohort/ case-control	Categorical SUA level	5	4,097	894	RR	2.83 (2.13, 3.76)	2.91E-12	78 (23, 89)	0.94	0.93	1.05-7.62	IV
Diabetic macrovascular complications (279)	T2DM patients	Case-control	Categorical SUA level	3	2,538	187	OR	1.03 (1.00, 1.06)	0.05	48 (0, 79)	0.45	0.01	0.56-2.30	IV
Diabetic retinopathy† (279)	T2DM patients	Case-control	Categorical SUA level	2	1,739	311	OR	1.23 (0.81, 1.87)	0.34	NA	NA	NP	NA	NS
Kidney disorders														
CKD incidence (282)	Middle-aged populations	Prospective/ retrospective cohort	1 mg/dL SUA increase	15	99,205	3,492	RR	1.22 (1.16, 1.28)	1.61E-14	66 (39, 78)	0.22	0.12	1.02-1.44	II
CKD new-onset incidence (283)	Non-CKD population	Prospective/ retrospective cohort	1 mg/dL SUA increase	7	153,620	7,014	HR	1.13 (1.04, 1.22)	2.74E-03	83 (63, 90)	0.12	0.24	0.88-1.44	IV
CKD new-onset incidence† (283)	Diabetic patients	Prospective/ retrospective cohort	Hyper vs normal	2	NA	NA	HR	1.90 (1.30, 2.78)	9.76E-04	NA	NA	0.94	NA	IV
eGFR (284)	Renal transplant recipients	Prospective/ retrospective cohort	Hyper vs normal	8	2,075	NA	MD to OR	0.36 (0.26, 0.52)	1.48E-08	66 (3, 82)	0.35	0.81	0.13-1.06	IV
SCr (284)	Renal transplant recipients	Prospective/ retrospective cohort	Hyper vs normal	5	873	NA	MD to OR	2.45 (1.69, 3.54)	2.77E-06	40 (0, 77)	0.15	0.65	0.88-6.81	IV
Graft loss (284)	Renal transplant recipients	Prospective/ retrospective cohort	Hyper vs normal	3	910	154	OR	2.28 (1.54, 3.38)	4.66E-05	0 (0, 73)	0.57	NP	0.18-29.36	IV

Outcome	Population	Study design included in MA	Comparison	N studies	N participants	N cases	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval	Evidence class[§]
Chronic allograft nephropathy (284)	Renal transplant recipients	Prospective/ retrospective cohort	Hyper vs normal	4	1,057	113	OR	2.81 (1.65, 4.77)	1.52E-04	26 (0, 75)	0.92	NP	0.53-14.76	IV
Neurocognitive disorders														
Alzheimer's disease (AD) (285)	General	Cohort/ case-control	SUA level (mg/dL)	21	3,617	1,128	MD to OR	0.29 (0.11, 0.76)	0.012	97 (96, 97)	0.30	NP	0.01-8.97	IV
Dementia/cognitive impairment (286)	General	Cohort/ case-control	SUA level (mg/dL)	31	7,021	2,681	SMD to OR	0.58 (0.41, 0.83)	2.52E-03	89 (86, 91)	0.01	0.004	0.08-4.48	IV
Vascular dementia (VaD) (286)	VaD patients vs controls	Cohort/ case-control	SUA level (mg/dL)	7	597	272	SMD to OR	0.92 (0.20, 4.12)	0.92	94 (90, 96)	0.45	<0.001	0.01-200.17	NS
Mild cognitive impairment (MCI) (286)	MCI patients vs controls	Cohort/ case-control	SUA level (mg/dL)	4	731	515	SMD to OR	0.65 (0.20, 2.17)	0.49	92 (83, 95)	0.36	0.52	0.01-63.36	NS
Parkinson's disease incidence (287)	General	Cohort and nested case-control	Hyper vs normal	6	33,185	578	RR	0.65 (0.43, 0.97)	0.04	42 (0, 73)	0.39	NP	0.24-1.77	IV
Multiple sclerosis (MS) (288)	MS patients vs control	Case-control	SUA level (µmol/L)	10	2,216	1,308	SMD to OR	0.49 (0.27, 0.87)	0.02	92 (88, 94)	0.11	NP	0.05-4.96	IV
Neruomyelistsoptic is (NMO) (288)	NMO patients vs control	Case-control	SUA level (µmol/L)	3	1,137	229	SMD to OR	0.22 (0.10, 0.45)	9.07E-05	82 (49, 91)	0.65	0.93	0.02-3.14	IV
Amyotrophic lateral sclerosis (ALS) (289)	ALS patients vs controls	Case-control	SUA level (mg/dL)	3	826	311	Hedge's G to OR	0.21 (0.14, 0.32)	6.33E-13	51 (0, 82)	0.43	NP	0.04-1.05	IV
Schizophrenia (Chronic) † (290)	Chronic Schizophrenia patients vs controls	Case-control	SUA level (mg/dL)	2	274	155	Hedge's G to OR	0.72 (0.43, 1.21)	0.22	NA	NA	NP	NA	NS

Outcome	Population	Study design included in MA	Comparison	N studies	N participants	N cases	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval	Evidence class§
Schizophrenia (First-Episode Psychosis) (290)	Schizophrenia patients vs controls	Case-control	SUA level (mg/dL)	3	277	103	Hedge's G to OR	0.37 (0.23, 0.59)	4.16E-05	0 (0, 73)	0.50	0.21	0.02-7.75	IV
Bipolar disorder (291)	Bipolar disorder vs controls	Case-control	SUA level (mg/dL)	9	1,127	619	SMD to OR	3.23 (1.82, 5.73)	7.09E-05	83 (66, 89)	0.19	NP	0.65-12.39	IV
Cancer outcomes														
Cancer incidence (292)	General	Prospective cohort	Highest vs lowest SUA category	5	456,053	14,355	RR	1.04 (0.99, 1.08)	0.08	45 (0, 78)	0.30	0.16	0.93-1.14	NS
Cancer incidence in digestive organs (292)	General	Prospective cohort	Highest vs lowest SUA category	3	266,347	2,521	RR	1.06 (0.96, 1.18)	0.27	53 (0, 79)	0.58	0.65	0.81-1.40	NS
Cancer in lymphoid and hematopoietic systems incidence† (292)	General	Prospective cohort	Highest vs lowest SUA category	2	86,739	397	RR	1.39 (1.13, 1.71)	1.89E-03	NA	NA	NP	NA	IV
Cancer incidence in male genital organs (292)	General	Prospective cohort	Highest vs lowest SUA category	3	162,022	2,634	RR	1.08 (0.96, 1.21)	0.19	61 (0, 87)	0.45	0.63	0.28-4.18	NS
Cancer incidence in respiratory system (292)	General	Prospective cohort	Highest vs lowest SUA category	4	456,053	2,941	RR	1.05 (0.93, 1.18)	0.43	71 (0, 87)	0.62	0.49	0.72-1.54	NS
Cancer in urinary organs incidence† (292)	General	Prospective cohort	Highest vs lowest SUA category	2	86,739	536	RR	1.17 (0.44, 3.15)	0.77	NA	NA	0.02	NA	NS

Outcome	Population	Study design included in MA	Comparison	N studies	N participants	N cases	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval	Evidence class[§]
All-cause and cause-specific mortality														
CHD mortality[*] (16)	General	Prospective cohort	Hyper vs normal	13	876,584	24,198	aRR	1.27 (1.16, 1.39)	3.46E-07	65 (36, 78)	0.10	NP	0.96-1.69	II
CVD mortality (293)	General	Prospective cohort	Highest vs lowest SUA category	9	165,806	6,121	RR	1.37 (1.19, 1.57)	1.07E-05	54 (0, 74)	0.59	NP	0.92-2.03	III
CVD mortality† (271)	Heart failure patients	Prospective cohort	Hyper vs normal	2	2,250	NA	HR	1.45 (1.18, 1.78)	4.25E-04	NA	NA	NP	NA	IV
CVD mortality† (270)	Hypertensive patients	Prospective/ retrospective cohort	Hyper vs normal	3	NA	NA	aHR	1.31 (0.96, 1.78)	0.09	NA	NA	NA	NA	NS
Stroke mortality (276)	General	Prospective cohort	Highest vs lowest SUA category	9	1,017,810	21,281	aRR	1.32 (1.23, 1.41)	1.11E-14	30 (0, 65)	0.92	NP	1.13-1.56	I*
CKD mortality (294)	General	Prospective cohort	1 mg/dL SUA increase	21	23,443	3,904	aHR	1.07 (1.04, 1.11)	5.46E-05	82 (74, 87)	0.04	0.03	0.93-1.24	III
Cancer mortality (292)	General	Prospective cohort	Highest vs lowest SUA category	12	632,472	NA	RR	1.17 (1.04, 1.32)	9.82E-03	66 (25, 80)	0.36	NP	0.82-1.69	IV
Cancer mortality in digestive organs (292)	General	Prospective cohort	Highest vs lowest SUA category	4	187,886	855	RR	1.22 (0.86, 1.74)	0.27	55 (0, 80)	0.99	NP	0.45-3.31	NS
Cancer mortality in bone, connective tissue, soft tissue, and skin† (292)	General	Prospective cohort	Highest vs lowest SUA category	NA	112,296	NA	RR	0.94 (0.47, 1.87)	0.87	NA	NA	NA	NA	NS

Outcome	Population	Study design included in MA	Comparison	N studies	N participants	N cases	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval	Evidence class[§]
Cancer mortality in lymphoid and hematopoietic systems† (292)	General	Prospective cohort	Highest vs lowest SUA category	NA	112,296	NA	RR	1.18 (0.82, 1.70)	0.38	NA	NA	NA	NA	NS
Cancer mortality in male genital organs† (292)	General	Prospective cohort	Highest vs lowest SUA category	NA	88,033	NA	RR	0.51 (0.07, 3.85)	0.52	NA	NA	NA	NA	NS
Cancer mortality in respiratory system and intrathoracic organs† (292)	General	Prospective cohort	Highest vs lowest SUA category	2	116,646	164	RR	1.08 (0.61, 1.91)	0.80	NA	NA	NP	NA	NS
Cancer mortality in urinary organs† (292)	General	Prospective cohort	Highest vs lowest SUA category	2	112,296	NA	RR	1.35 (0.88, 2.07)	0.17	NA	NA	NP	NA	NS
All-cause mortality (271)	Heart failure patients	Cohort/ case-control	Hyper vs normal	11	12,444	1,888	HR	2.15 (1.64, 2.83)	6.64E-08	81 (67, 88)	0.01	0.37	0.87-5.31	II
Short-term mortality(275)	AMI patients	Prospective/ retrospective cohort	Highest vs lowest SUA category	8	6,805	396	OR	3.24 (2.47, 4.27)	3.75E-16	31 (0, 69)	0.83	NP	1.74-6.06	IV
Medium term mortality (275)	AMI patients	Prospective/ retrospective cohort	Highest vs lowest SUA category	5	5,194	565	OR	2.69 (2.00, 3.62)	1.75E-10	55 (0, 81)	0.66	NP	1.09-6.67	IV
In-Hospital mortality (296)	AMI patients	Cohort	Hyper vs normal	6	5,686	218	RR	2.10 (1.03, 4.26)	0.04	81 (51, 90)	0.86	NP	0.21-20.66	IV
All-cause mortality (279)	T2DM patients	Cohort/ case-control	Hyper vs normal	3	5,534	NA	HR	1.09 (1.03, 1.17)	8.04E-03	19 (0, 73)	0.49	NP	0.90-1.33	IV
All-cause mortality (293)	General	Prospective cohort	Highest vs lowest SUA category	10	143,483	7,031	RR	1.23 (1.08, 1.39)	1.35E-03	75 (56, 84)	0.51	NP	0.79-1.90	IV

Outcome	Population	Study design included in MA	Comparison	N studies	N participants	N cases	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval	Evidence class[§]
All-cause mortality (274)	Patients after PCI	Prospective/ retrospective cohort	Hyper vs normal	9	17,268	NA	RR	1.52 (1.28, 1.81)	2.95E-06	64 (3, 81)	0.002	NP	0.98-2.24	IV
All-cause mortality (270)	Hypertensive patients	Prospective/ retrospective cohort	Hyper vs normal	4	46,103	5,820	aHR	1.12 (1.02, 1.23)	0.02	26 (0, 76)	0.77	0.93	0.86-1.49	IV
All-cause mortality (270)	CKD population	Prospective/ retrospective cohort	Hyper vs normal	5	1,789	609	RR	1.67 (1.29, 2.16)	1.09E-04	NA	NA	NA	NA	IV
Other outcomes														
Medium/long-term occurrence of death/MACE (275)	AMI patients	Prospective/ retrospective cohort	50 µmol/L increase	4	3,533	NA	aHR	1.19 (1.03, 1.37)	0.02	84 (47, 92)	0.03	NP	0.65-2.18	IV
Short-term occurrence of death/MACE (275)	AMI patients	Prospective/ retrospective cohort	Highest vs lowest SUA category	4	3,625	336	aOR	2.26 (1.85, 2.77)	1.61E-14	0 (0, 68)	0.97	0.23	1.45-3.53	IV
Combined death or cardiac events (271)	Heart failure patients	Cohort, case-control and Post-hoc RCT	Hyper vs normal	9	12,699	1,765	HR	1.39 (1.18, 1.63)	7.44E-05	66 (13, 82)	0.001	0.12	0.89-2.07	III
Adverse outcomes (mortality, MACE, In-stent restenosis) (274)	Patients after PCI	Prospective/ retrospective cohort	Hyper vs normal	12	21,030	NA	RR	1.46 (1.29, 1.65)	3.63E-09	59 (3, 77)	<0.001	NP	1.05-1.95	IV
Occurrence of poor outcomes (297)	Acute ischaemic stroke patients	Prospective/ retrospective cohort, or nested case–control	Highest vs lowest SUA category	9	7,932	NA	HR	0.77 (0.68, 0.88)	8.12E-05	44 (0, 73)	0.30	NP	0.56-1.06	IV
Psoriasis (298)	Psoriasis vs controls	Case-control	SUA level (mg/dL)	13	29,037	1,644	MD to OR	4.46 (1.57, 12.62)	4.96E-03	98 (98, 99)	0.41	<0.001	0.06-320.30	IV

Outcome	Population	Study design included in MA	Comparison	N studies	N participants	N cases	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval	Evidence class[§]
Severe psoriasis (298)	Sever psoriasis patients vs controls	Case-control	SUA level (mg/dL)	3	300	104	MD to OR	1.57 (0.25, 9.80)	0.64	92 (78, 96)	0.20	<0.001	0.00-1.52E-10	NS
Non-alcoholic fatty liver disease (NAFLD) (299)	General	Prospective/ retrospective cohort, or case–control	Highest vs lowest SUA category	9	55,573	10,581	OR	1.92 (1.59, 2.31)	2.51E-11	78 (61, 86)	0.02	NP	0.99-3.74	II

Abbreviations: MA, meta-analysis; AF, atrial fibrillation; CHD, coronary heart disease; CVD, cardiovascular disease; MACE, major adverse cardiovascular events; AMI, acute myocardial infarction; PCI, percutaneous coronary intervention; T2DM, Type 2 diabetes; CKD, chronic kidney disease; AD, Alzheimer's disease; PD, Parkinson's disease; VaD, Vascular dementia; MCI, mild cognitive impairment; MS, multiple sclerosis; NMO, neruomyelistsopticis; ALS, amyotrophic lateral sclerosis; NAFLD, non-alcoholic fatty liver disease; MD, mean difference; SMD, standard mean difference; NA, not available; NS, not significant; NP, not pertinent (because the number of expected significant studies was larger than the number of observed significant studies).

† The heterogeneity (I^2), Egger's test or 95% PI could not be calculated, either because data about the individual component studies were insufficient or because the number of studies included in meta-analyses was less than 3.

§ Evidence class criteria: (i) Class I (convincing): statistical significance with $p<10^{-6}$, more than 1,000 cases (or more than 20,000 participants for continuous outcomes), the largest component study reported statistically significant effect ($p<0.05$); 95% prediction interval excluded the null; no large heterogeneity ($I^2<50\%$), no evidence of small study effects ($p>0.10$) and excess significance bias ($p>0.10$); (ii) Class II (highly suggestive): statistical significance with $p<10^{-6}$, more than 1,000 cases (or more than 20,000 participants for continuous outcomes), the largest component study reported statistically significant effect ($p<0.05$); (iii) Class III (suggestive): statistical significance with $p<10^{-3}$, more than 1,000 cases (or more than 20,000 participants for continuous outcomes); (iv) Class IV (weak): the remaining statistically significant associations with $p<0.05$.

*Evidence was re-assessed by examining the meta-analyses in depth to verify the eligibility/appropriateness of the data included in analysis and errors were found; when errors and analyses were corrected, the association became non-statistically significant.

Table 3 - 2: Re-assessing the credibility of associations with class I-II evidence reported in meta-analyses of observational studies*

Outcome	Population	Study design included in MA	Comparison	N studies	N participants	N cases	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval	Evidence class§
Stroke mortality	General	Prospective cohort	Highest vs lowest SUA category	8	600,076	5,205	aRR	1.17 (0.91, 1.51)	0.22	84 (73, 89)	0.44	NP	0.46-2.98	NS (changed from I)
Heart failure incidence	General	Prospective cohort	Hyper vs. normal	5	427,917	10,171	HR	1.65 (1.41, 1.94)	1.77E-09	72 (7, 86)	0.49	0.31	1.05-2.61	II
Hypertension incidence	General	Prospective cohort	Hyper vs. normal	12	68,401	16,132	aRR	1.42 (1.27, 1.59)	2.16E-09	76 (53, 85)	0.04	NP	0.98-2.05	II
IFG/T2DM	General	Prospective cohort	Highest vs lowest SUA category	13	56,130	5,629	RR	1.62 (1.47, 1.77)	1.25E-22	0 (0, 49)	0.07	NP	1.45-1.79	II
CKD incidence	Middle-Aged Populations	Prospective cohort	1 mg/dL SUA increase	12	78,205	2,793	RR	1.19 (1.12, 1.25)	1.26E-09	67 (34, 80)	0.10	0.15	0.99-1.42	II
CHD mortality	General	Prospective cohort	Hyper vs normal	13	876,584	24,198	aRR	1.27 (1.16, 1.39)	3.47E-07	65 (36, 78)	0.10	NP	0.96-1.69	II
All-cause mortality	HF patients	Prospective cohort	Hyper vs. normal	6	9,608	1,474	HR	2.38 (1.59, 3.56)	2.98E-05	88 (77, 92)	0.05	0.39	0.61-9.35	III (changed from II)
Non-alcoholic Fatty Liver Disease (NAFLD)†	General	Prospective cohort	Highest vs lowest SUA category	2	12,631	2,530	OR	1.43 (1.20, 1.71)	8.63E-05	NA	NA	NP	NA	III (changed from II)

Abbreviations: MA, meta-analysis; AF, atrial fibrillation; CHD, coronary heart disease; T2DM, Type 2 diabetes; NA, not available; NS, not significant; NP, not pertinent (because the number of expected significant studies was larger than the number of observed significant studies).

† The heterogeneity (I^2), Egger's test and 95% PI could not be calculated, because the number of studies included in meta-analyses was less than 3.

§ Evidence class criteria: (i) Class I (convincing): statistical significance with $p<10^{-6}$, more than 1,000 cases (or more than 20,000 participants for continuous outcomes), the largest component study reported statistically significant effect ($p<0.05$); 95% prediction interval excluded the null value; no large heterogeneity ($I^2<50\%$), no evidence of small study effects ($p>0.10$) and excess significance bias ($p>0.10$); (ii) Class II (highly suggestive): statistical significance with $p<10^{-6}$, more than 1,000 cases (or more than 20,000 participants for continuous outcomes), the largest component study reported statistically significant effect ($p<0.05$); (iii)

Class III (suggestive): statistical significance with $p<10^{-3}$, more than 1,000 cases (or more than 20,000 participants for continuous outcomes); (iv) Class IV (weak): the remaining statistically significant associations with $p<0.05$.

*Evidence was re-assessed by examining the meta-analyses in depth to verify the eligibility/appropriateness of the data included in analysis or excluding the data from retrospective/case-control studies to address reverse causality.

3.4.3 Meta-analyses of randomised controlled trials

We identified 31 meta-analyses of RCTs on SUA-lowering therapy from 8 publications (**Supplementary Table 3-4**). The median number of studies included in the meta-analyses was 5 (range: 2-10) and the median number of participants was 216 (range: 41-738). More than one meta-analysis was found for 5 outcomes (**Supplementary Table 3-4**). The direction and statistical significance of the effects in overlapping meta-analyses were in concordance only for one (20.0%) outcome: serum creatinine (SCr) (n=2) (316, 317). Discordance in either the direction and/or the statistical significance was found for the remaining 4 outcomes: glomerular filtration rate (eGFR) (n=2) (316, 317), end-stage kidney disease (n=2) (316, 317), systolic blood pressure (SBP) (n=2) (317, 320), and diastolic blood pressure (DBP) (n=2) (317, 320).

Twenty unique meta-analyses (**Table 3-3**) were identified for the outcomes in relation to kidney disorders (n=10), endothelial function (n=2), all-cause and cause-specific mortality (n=4), and other outcomes (n=4). In **Supplementary Figure 3-3** we plot the summary effects of the unique meta-analyses of RCTs. Overall, 1^2 (60.0%) of unique meta-analyses of RCTs reported a nominally statistically significant summary result at $p<0.05$ (8 had $p<0.001$). Only 3 (15.0%) meta-analyses had a 95% PI that excluded the null (two nephrolithiasis outcomes [with thiazide and citrate treatment] and one renal function outcome); 11 (55.0%) meta-analyses showed no large heterogeneity ($I^2<50\%$); and 13 (65.0%) meta-analyses showed neither small study effects nor excess significant bias.

Taken all these together, only one outcome (recurrence of nephrolithiasis [with citrates treatment]) reported a p<10-3, had a 95% PI excluding the null and had no evidence of large heterogeneity or bias. In the original meta-analyses, the strength of evidence was graded collectively for three nephrolithiasis outcomes (with thiazide, citrate and allopurinol treatment) by using an approach conceptually similar to the GRADE ranking system (361) and evidence for these three nephrolithiasis outcomes was graded as moderate.

Table 3 - 3: Health outcomes reported in meta-analysis of RCTs.

Outcome	Population	SUA-lowering therapy	N studies	N participants	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval
Kidney disorders											
Recurrence of nephrolithiasis[†§] (314)	Nephrolithiasis patients	Allopurinol	2	152	RR	0.59 (0.42, 0.84)	2.90E-03	NA	NA	0.39	NA
Recurrence of nephrolithiasis[§] (314)	Nephrolithiasis patients	Thiazides	5	300	RR	0.52 (0.39, 0.69)	9.00E-06	0 (0, 64)	0.06	0.11	0.33-0.82
Recurrence of nephrolithiasis[§] (314)	Nephrolithiasis patients	Citrates	4	197	RR	0.26 (0.15, 0.45)	2.84E-06	0 (0, 68)	0.19	NP	0.08-0.88
SCr (315)	General	All active therapy	9	580	SMD to OR	0.10 (0.03, 0.39)	4.64E-04	93 (90, 95)	0.39	NP	0.01-13.21
SCr (316)	CKD Patients	Allopurinol	6	354	MD to OR	0.16 (0.08, 0.34)	1.00E-06	70 (0, 85)	0.01	0.59	0.02-1.76
eGFR (315)	General	All active therapy	3	218	SMD to OR	2.22 (1.21, 4.06)	9.79E-03	29 (0, 80)	0.24	NP	0.01-497.40
eGFR (317)	Patients with CKD or decreased kidney function	Allopurinol	5	346	MD to OR	1.18 (0.97, 1.42)	0.09	0 (0, 64)	0.29	NP	0.86-1.60
Proteinuria (317)	Patients with CKD or decreased kidney function	Allopurinol	5	250	MD to OR	0.91 (0.73, 1.12)	0.40	0 (0, 64)	0.42	NP	0.64-1.28
Blood urea nitrogen (BUN) (316)	CKD patients	Allopurinol	3	169	MD to OR	0.18 (0.10, 0.32)	1.47E-08	0 (0, 73)	0.88	0.67	0.01-7.16
End-stage renal disease (316)	CKD patients	Allopurinol	5	267	RR	0.33 (0.21, 0.51)	1.38E-06	0 (0, 64)	0.01	0.07	0.16-0.68
Endothelial function											
Flow-mediated dilatation (318)	Population with vascular disease/risk factors	Allopurinol/ oxypurinol	5	144	MD to OR	4.38 (1.85, 10.38)	8.76E-04	60 (0, 83)	0.23	0.24	0.27-70.69
Forearm blood flow (318)	Population with vascular disease/risk factors	Allopurinol/ oxypurinol	5	148	MD to OR	2.69 (1.22, 5.93)	0.014	53 (0, 81)	0.09	0.61	0.24-30.73

Outcome	Population	SUA-lowering therapy	N studies	N participants	Type of metric	Relative risk (95%CI)	P-value	I^2 (95%CI)	P-value for Egger test	P-value for excess significance test	95% prediction interval
Mortality											
Death during neonatal or infancy (319)	All infants	Allopurinol	3	114	RR	0.87 (0.43, 1.75)	0.71	34 (0, 81)	0.49	NP	0.01-952.4
Death during neonatal or infancy† (319)	Infants with severe hypoxic-ischaemic encephalopathy	Allopurinol	2	41	RR	0.92 (0.39, 2.15)	0.86	NA	NA	NP	NA
Death or serve neurodevelopmental disability (319)	All infants	Allopurinol	3	110	RR	0.85 (0.63, 1.15)	0.29	0 (0, 73)	0.12	NP	0.12-5.98
Death or serve neurodevelopmental disability† (319)	Infants with severe hypoxic-ischaemic encephalopathy	Allopurinol	2	41	RR	0.93 (0.67, 1.30)	0.68	NA	NA	NP	NA
Others outcomes											
Severe quadriplegia (319)	Surviving infants with hypoxic-ischaemic encephalopathy	Allopurinol	3	73	RR	0.58 (0.27, 1.26)	0.17	0 (0, 73)	0.69	NP	0.01-86.99
Seizures in neonatal period (319)	Surviving infants with hypoxic-ischaemic encephalopathy	Allopurinol	3	114	RR	0.98 (0.84, 1.15)	0.81	0 (0, 73)	0.15	NP	0.35-2.79
SBP* (320)	Patients with elevated SUA or kidney dysfunction	Allopurinol	10	738	MD (mmHg)	-3.33 (-5.25, -1.42)	0.001	87 (79, 91)	0.60	NP	-13.61-6.94
DBP* (320)	Patients with elevated SUA or kidney dysfunction	Allopurinol	10	738	MD (mmHg)	-1.29 (-2.48, -0.10)	0.03	82 (68, 88)	0.38	NP	-8.22-5.65

Abbreviations: RR, relative risk; CKD, chronic kidney disease; SCr, serum creatinine; eGFR, glomerular filtration rate; SBP, systolic blood pressure; DBP, diastolic blood pressure; MD, mean difference; SMD, standard mean difference; NA, not applicable (did not calculate with only 2 studies); NS, not significant; NP, not pertinent (because the number of expected significant studies was larger than the number of observed significant studies).

§ The strength of evidence was graded based on the evidence-based practice centre approach (conceptually similar to the GRADE ranking system); recurrence of nephrolithiasis (with allopurinol, thiazides or citrates treatment) was all considered with moderate evidence in original meta-analyses.

† The heterogeneity (I2), Egger's test or 95% PI could not be calculated, because the number of studies included in meta-analyses was less than 3.

* Meta-analyses included one prospective study.

3.4.4 Mendelian randomisation studies

A total of 107 MR analyses were identified from 36 publications (**Supplementary Table 3-5**). The median number of participants and median number of cases were 7,158 (range: 343-206,822) and 2,225 (range: 19-65,877), respectively. The proportion of SUA variance (adjusted R^2) explained by genetic instruments was 1.8%-6.0%. More than one MR study was identified for 14 outcomes (**Supplementary Table 3-5**). Discordance in either the direction and/or the statistical significance of association among overlapping MR existed for all the identified outcomes: BMI (n=7) (322, 323, 328, 329, 337, 342, 348), BMD in femoral neck (n=2) (324, 325), CHD (n=5) (323, 327, 333, 345, 353), DBP (n=7) (323, 328, 333, 337, 346, 348, 351), SBP (n=7) (323, 328, 333, 337, 346, 348, 351), metabolic syndrome (n=2) (334, 347), glucose (n=3) (323, 333, 348), TG (n=3) (323, 348, 350), diabetes (n=6) (323, 326, 332), (327, 349, 354), SCr (n=2) (337, 356), eGFR (n=5) (333, 337, 348, 355, 356), Parkinson's disease (n=5) (338, 339, 343, 344, 352), memory performance (n=2) (341), and gout (n=3) (326, 327, 333).

The 56 unique outcomes (**Table 3-4**) investigated in individual MR studies belonged to the following categories: anthropometric variables (n=9), cardiovascular outcomes (n=15), kidney disorders (n=6), metabolic disorders (n=5), neurocognitive disorders (n=5), metabolites (n=11), all-cause and cause-specific mortality (n=3) and other outcomes (n=2). Only 9 (16.1%) outcomes (diabetic macrovascular disease, arterial stiffness [internal diameter of carotid artery], adverse renal events, Parkinson's disease, lifetime anxiety disorders, memory performance, CVD mortality, sudden cardiac death, and gout) presented statistically significant associations of $p<0.05$. Three MR studies (on memory performance, Parkinson's disease and gout) reported discordant results in the direction and/or statistical significance in other MR studies. Of note, only 4 outcomes (diabetic macrovascular disease, arterial stiffness [internal diameter of carotid artery], renal events, and gout) reported a $p<0.01$, and only that for gout was based on convincing evidence ($p=3.55\times10^{-40}$, $N_{\text{sample size}}=71,501$, power>99%).

3.4.5 Comparing findings from meta-analyses

Outcomes reported in meta-analyses of observational studies with highly suggestive evidence or meta-analyses of RCTs with 95% PI excluding the null are summarised in **Table 3-5**. Among these outcomes, hypertension and CKD showed concordant evidence between meta-analyses of observational studies and the selected (largest) meta-analyses of RCTs on their corresponding intermediate traits or surrogate outcomes (e.g. SBP, DBP, SCr, eGFR

and end-stage renal disease), but had discordant evidence from MR studies. Moreover, as we mention above, even for these outcomes, there were additional meta-analyses of RCTs that had found discordant effects in terms of direction and/or statistical significance for all these intermediate traits or surrogate outcomes with the exception of SCr. Heart failure, impaired fasting glucose or diabetes and CHD mortality showed no evidence from meta-analyses of RCTs, and MR studies reported discordant evidence on either the corresponding outcomes, intermediate traits or surrogate outcomes. Recurrence of nephrolithiasis was only reported in meta-analysis of RCTs, and no evidence was found from meta-analyses of observational studies or MR studies.

Table 3 - 4: Health outcomes reported in Mendelian randomisation studies.

Outcomes	Population	N/n Events (N studies)*	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate of effect (95%CI)	P-value	Statistical power#
Anthropometric variables								
Appendicular lean mass (kg) (321)	British	3,953	rs737267 in *SCL2A9*	NA	β	0.013 (NA, NA)	0.51	NA
Fat mass (kg) (322)	Swiss	6,184	rs6855911 in *SCL2A9*	3.2%	β	0.05 (-0.10, 0.19)	0.52	0.07
BMI (kg/m^2) (323)	European	127,600 (64)*	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	-0.0003 (-0.0008, 0.0002)	NA	NA
Waist circumference (cm) (322)	Swiss	6,184	rs6855911 in *SCL2A9*	3.2%	β	0.08 (-0.05, 0.21)	0.24	0.06
BMD in femoral neck(g/cm^2) (324)	American	2,501	Genetic risk score of 5 SUA-related SNPs	3.3%	β	-0.27 (-0.58, 0.03)	0.08	0.07
BMD in L1–L4 (g/cm^2) (325)	Chinese	1,667	Genetic risk score of 5 SUA-related SNPs	1.8%	β	0.39 (-0.26, 0.98)	0.26	0.19
BMD in spine (g/cm^2) (324)	American	2,501	Genetic risk score of 5 SUA-related SNPs	3.3%	β	0.08 (-0.32, 0.48)	0.68	0.18
BMD in total femur (g/cm^2) (324)	American	2,501	Genetic risk score of 5 SUA-related SNPs	3.3%	β	-0.29 (-0.60, 0.01)	0.06	0.11
BMD in total hip (g/cm^2) (325)	Chinese	1,667	Genetic risk score of 5 SUA-related SNPs	1.8%	β	0.19 (-0.36, 0.74)	0.50	0.19
Cardiovascular outcomes								
Arrhythmia (326)	German	3,060/444	Genetic risk score of 8 SUA-related SNPs	NA	OR	0.98 (0.88, 1.08)	0.64	0.05‡
Atrial fibrillation (326)	German	3,060/368	Genetic risk score of 8 SUA-related SNPs	NA	OR	1.03 (0.93, 1.15)	0.57	0.05‡
Cardiomyopathy (326)	German	3,060/316	Genetic risk score of 8 SUA-related SNPs	NA	OR	1.00 (0.89, 1.12)	0.93	0.05‡
CHD (323)	European	206,822/65,877 (58)*	Genetic risk score of 31 SUA-related SNPs	4.2%	OR	1.05 (0.92, 1.18)	0.49	0.57
Heart failure (327)	Pakistani	22,926/4,526 (2)*	Genetic risk score of 14 SUA-related SNPs	3.1%	OR	1.07 (0.88, 1.30)	0.51	0.11
Ischaemic heart disease (328)	Danish	68,674/3,742 (2)*	rs7442295 in *SCL2A9*	2.2%	HR	0.93 (0.79, 1.09)	0.38	0.10
Hypertension (326)	German	3,060/2,225	Genetic risk score of 8 SUA-related SNPs	NA	OR	0.98 (0.90, 1.06)	0.56	0.05‡
Ischaemic stroke (327)	Pakistani	82,091/14,779 (2)*	Genetic risk score of 14 SUA-related SNPs	3.1%	OR	0.99 (0.88, 1.12)	0.93	0.05
Peripheral vascular disease (326)	German	3,060/295	Genetic risk score of 8 SUA-related SNPs	NA	OR	0.92 (0.82, 1.04)	0.18	0.06‡

Outcomes	Population	N/n Events (N studies)*	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate of effect (95%CI)	P-value	Statistical power#
Valve disease (326)	German	3,060/538	Genetic risk score of 8 SUA-related SNPs	NA	OR	1.08 (0.99, 1.19)	0.10	0.07‡
Diabetic macrovascular disease (330)	Chinese T2DM patients	3,207	Genetic risk score of 3 SUA-related SNPs	NA	OR	1.18 (1.06, 1.33)	0.004	NA
cIMT (mm) (329)	Finnish (male)	1,985	rs13129697 in *SCL2A9*	NA	β	<0.0001 (NA, NA)	0.99	NA
Arterial stiffness (internal diameter of carotid artery) (mm) (331)	Italian	449	rs734553 in *SLC2A9*	NA	β	0.48 (NA, NA)	0.003	NA
DBP (mm Hg) (323)	European	89,667 (37)*	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	0.005 (0.003, 0.007)	NA	NA
SBP (mm Hg) (323)	European	89,667 (37)*	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	0.005 (0.003, 0.006)	NA	NA
Metabolic disorders								
T2DM (327)	Pakistani	110,452/26,488 (2)*	Genetic risk score of 14 SUA-related SNPs	3.1%	OR	0.95 (0.86, 1.05)	0.28	0.24
Diabetes (332)	European	165,482/41,508 (2)*	Genetic risk score of 24 SUA-related SNPs	4.0%	OR	0.99 (0.92, 1.06)	0.79	0.06
Fasting glucose (mmol/L) (323)	European	57,397 (28)*	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	-0.001 (-0.003, 0.001)	NA	NA
Fasting insulin† (333)	American	19,899 (5)*	Genetic risk score of 8 SUA-related SNPs	6.0%	Z statistic	-0.015 (NA, NA)	0.99	NA
Metabolic syndrome (334)	Chinese	7,827	Genetic risk score of 2 SNPs (rs11722228 in *SLC2A9* and rs2231142 in *ABCG2*)	2.1%	OR	1.03 (0.98, 1.09)	0.23	NA
Kidney disorders								
CKD (333)	American	23,387/3,092 (5)*	Genetic risk score of 8 SUA-related SNPs	6.0%	OR	1.20 (0.96, 1.50)	0.12	0.70
Acute kidney injury (335)	American	7,553/823	Genetic risk score of 8 SUA-related SNPs	6.0%	HR	1.01 (0.77, 1.34)	0.92	0.05
Adverse renal events (336)	Italian	755/244	rs734553 in *GLUT9*	NA	HR	2.35 (1.25, 4.42)	0.01	NA
Log eGFR (mL/min/1.73 m^2) (333)	American	23,844 (5)*	Genetic risk score of 8 SUA-related SNPs	6.0%	β	0.001 (-0.01, 0.02)	0.91	0.05
SCr (mmol/L) (337)	European	7,979 (2)*	Genetic risk score of 5 SUA-related SNPs	2.3%	β	-19.23 (-40.32, 1.86)	0.07	NA
Albumin/creatinine ratio (356)	Native American	3,604 (3)*	Genetic risk score of 5 SUA-related SNPs	5.3%	Residual variance¶	Overall p>0.05		NA

Outcomes	Population	N/n Events (N studies)*	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate of effect (95%CI)	P-value	Statistical power#
Neurocognitive disorders								
Parkinson's disease (PD) (343)	Birtish	1,815/1,061	Genetic risk score of 8 SUA-related SNPs	NA	OR	1.55 (1.10, 2.18)	0.01	0.59‡
Age at onset of PD (339)	European	664 (3)*	4 SNPs in *SCL2A9*	NA	β	Null after multiple testing correction		
			rs737267	NA		3.10 (0.17, 6.03)	0.04	NA
			rs6449213	NA		-1.18 (-4.96, 2.59)	0.54	
			rs1014290	NA		-4.56 (-8.13, -1.00)	0.01	
			rs733175	NA		3.59 (0.67, 6.51)	0.02	
Lifetime anxiety disorders (340)	Swiss	3,716	rs6855911 in *SLC2A9*	3.2%	OR (male)	1.40 (1.07, 1.84)	0.02	0.11
					OR (female)	0.97 (0.80, 1.17)	0.73	0.05
Current anxiety disorders (340)	Swiss	3,716	rs6855911 in *SLC2A9*	3.2%	OR (male)	1.42 (0.99, 2.03)	0.06	0.12
					OR (female)	0.84 (0.66, 1.06)	0.14	0.07
Memory performance (341)	European: Population 1	1,091	4 SNPs in *SCL2A9*	NA	β	Overall p<0.05		NA
	European: Population 2	1,066	4 SNPs in *SCL2A9*	NA	β	Overall p>0.05		NA
Metabolites								
HDL-C (mmol/L) (323)	European	196,621 (68)*	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	-0.008 (-0.010, -0.006)	NA	NA
LDL-C (mmol/L) (323)	European	196,621 (68)*	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	-0.001 (-0.003, 0.001)	NA	NA
TC (mmol/L) (323)	European	196,621 (68)*	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	0.000 (-0.002, 0.002)	NA	NA
TG (mmol/L) (323)	European	196,621 (68)*	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	0.01 (0.01, 0.02)	NA	NA
Parathyroid hormone (pg/mL) (325)	Chinese	1,667	Genetic risk score of 5 SUA-related SNPs	1.8%	β	-0.63 (-2.12, 0.85)	0.40	0.05
Phosphorus (mmol/L) (325)	Chinese	1,667	Genetic risk score of 5 SUA-related SNPs	1.8%	β	-0.16 (-0.74, 0.42)	0.59	0.05
CRP (mg/L) (342)	European	7,158	Genetic risk score of 29 SUA-related SNPs	NA	β	-0.05 (-0.15, 0.05)	0.37	NA
Calcium (mmol/L) (325)	Chinese	1,667	Genetic risk score of 5 SUA-related SNPs	1.8%	β	0.06 (-0.10, 0.21)	0.48	0.20

Outcomes	Population	N/n Events (N studies)*	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate of effect (95%CI)	P-value	Statistical power#
Tropocollagen type 1 N-terminal propeptide (ng/L) (325)	Chinese	1,667	Genetic risk score of 5 SUA-related SNPs	1.8%	β	0.11 (-1.53, 1.75)	0.90	0.05
β-crosslaps of type I collagen containing cross-linked C telopeptide (ng/L) (325)	Chinese	1,667	Genetic risk score of 5 SUA-related SNPs	1.8%	β	-1.45 (-3.17, 0.27)	0.10	0.05
25(OH)D (ng/mL) (325)	Chinese	1,667	Genetic risk score of 5 SUA-related SNPs	1.8%	β	0.76 (-0.63, 2.15)	0.28	0.05
All-cause and cause-specific mortality								
Cardiovascular mortality (326)	German	3,060/NA	Genetic risk score of 8 SUA-related SNPs	NA	aHR	1.11 (1.02, 1.21)	0.02	NA
All-cause mortality (326)	German	3,060/NA	Genetic risk score of 8 SUA-related SNPs	NA	aHR	1.02 (0.95, 1.09)	0.59	NA
Sudden cardiac death (326)	Germany	3,060/NA	Genetic risk score of 8 SUA-related SNPs	NA	aHR	1.18 (1.03, 1.35)	0.02	NA
Other outcomes								
Cancer (326)	German	3,060/226	Genetic risk score of 8 SUA-related SNPs	NA	OR	0.95 (0.83, 1.08)	0.41	0.05‡
Gout (327)	Pakistani	71,501/3,151 (2)*	Genetic risk score of 14 SUA-related SNPs	3.1%	OR	5.84 (4.56, 7.49)	3.55E-40	1.00

Abbreviations: GI: Genetic instruments; BMI, body mass index; BMD, bone mineral density; SBP, systolic blood pressure; DBP, diastolic blood pressure; CHD, coronary heart disease; IMT, intima-media thickness; CVD, cardiovascular disease; T2DM, Type 2 diabetes; CKD, chronic kidney disease; SCr, serum creatinine; eGFR, glomerular filtration rate; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TG, triglyceride; TC, total cholesterol; CRP, C-reactive protein; PD, Parkinson's disease; MD, mean difference; NA, not available.

* If the outcomes were reported from Mendelian randomisation analysis with pooling multiple studies, the number of studies included in pooled analysis was displayed in brackets.

† Because of the lack of a standard to covert insulin in different studies to the same scale, sample size-weighted pooled analysis were performed and Z statistics were reported instead of the β coefficient.

§ MD (mean difference) represented the difference in mean caused by per inverse variance weighted allele estimated from pooled analysis.

¶ Residual variance represented the proportion of residual variance explained by the SUA related SNPs.

‡ The statistical power was a crude estimation, as the MR studies failed to report R^2; we used the extrapolated R^2 from other MR studies that used the same genetic variants as instruments for calculation.

When MR studies that did not provide other necessary information for calculation (e.g. SD of SUA levels, SD of outcomes, or the number of cases), the statistical power was not calculated (reported as NA).

Table 3 - 5: Summary of evidence grading and comparison of multiple evidence.

Outcomes	MA of observational studies	MA of RCTs*	MR studies
Heart failure	Class II	NA	Heart failure: ($N_{sample\ size}$=22,926; p=0.51; power=0.11)
Hypertension†	Class II	SBP: (p=0.001; 95% PI included null) DBP: (p=0.034; 95% PI included null)	Hypertension: ($N_{sample\ size}$=3,060; p=0.56; power=0.05)
Impaired fasting glucose or diabetes	Class II	NA	Diabetes: ($N_{sample\ size}$=165,482; p=0.79; power=0.06) Fasting glucose: ($N_{sample\ size}$=57,397; p>0.05) Fasting insulin: ($N_{sample\ size}$=19,899; p=0.99)
CKD†	Class II	SCr: ($p=4.64\times10^{-4}$; 95% PI included null) eGFR: ($p=9.79\times10^{-3}$; 95% PI included null) End stage renal disease: ($p=1.38\times10^{-6}$; 95% PI excluded null)	CKD: ($N_{sample\ size}$=23,387; p=0.12; power=0.70) Adverse renal events: ($N_{sample\ size}$=755; p=0.01) SCr: ($N_{sample\ size}$=7,979; p=0.07) eGFR: ($N_{sample\ size}$=23,844; p=0.91;power=0.05)
CHD mortality†	Class II (general population)	NA	CHD: incidence ($N_{sample\ size}$=206,822; p=0.49; power=0.57)
Recurrence of nephrolithiasis	NA	Citrates treatment: ($p=2.84\times10^{-6}$; 95% PI excluded null) Thiazides treatment: ($p=9.00\times10^{-6}$; 95% PI excluded null)	NA

Abbreviations: MA, meta-analyses; MR, Mendelian randomisation; CKD, chronic kidney disease; CHD, coronary heart disease; HF, heart failure; NA, not applicable; NS, not significant. † If there was no identical outcomes investigated in meta-analyses of RCTs and/or MR studies to match with the class I-II observational associations, we juxtaposed the corresponding intermediate traits as surrogates for comparison.

*Data presented on the largest meta-analysis of RCTs for each outcome; for SBP, DBP, eGFR, and end-stage renal disease other meta–analyses of RCTs on the same outcomes showed discordant results in direction of effect and/or statistical significance.

3.5 Discussion

In this study, we provide a comprehensive overview of reported associations between SUA levels and a wide range of health outcomes by incorporating evidence from meta-analyses of observational studies, meta-analyses of RCTs and MR studies. We also further evaluated the reported evidence by following criteria that we have previously applied to appraise the epidemiological credibility in several research fields (255, 362, 363). In summary, our study comprised 76 unique meta-analyses of observational studies, 20 unique meta-analyses of RCTs and 56 unique individual MR studies, which overall covered 136 unique health outcomes.

3.5.1 Main findings and possible explanations

Most health outcomes which were reported to be associated with SUA were identified from meta-analyses of observational studies, but after the application of our criteria none of them was classified as convincing (class I). Highly suggestive evidence (class II) existed for 5 health outcomes, including heart failure, hypertension, impaired fasting glucose or diabetes, CKD and CHD mortality in general population. It is notable that a large proportion (80.3%) of the examined meta-analyses displayed substantial heterogeneity (I^2>50%), indicating that these associations should be interpreted with caution. Possible sources of the observed heterogeneity include the mixture of prospective, retrospective or case-control studies and the mixture of different comparison groups, since some meta-analyses synthesised individual studies with diverse contrasted categories of SUA levels (e.g. various choices of tertiles, quartiles, quintiles or sextiles of SUA levels). Likewise, although the outcomes with class I-II evidence fulfilled the criteria of credibility assessment for meta-analyses of observational studies, it would be inadvisable to conclude causation on this basis alone, due to the inherent limitations of unmeasured confounding, undetected bias or reverse causality in observational studies. In relation to reverse causality for example, some of the associations which were initially classified as class II (e.g. all-cause mortality in heart failure patients and non-alcoholic fatty liver disease), were no longer highly suggestive (and were downgraded to class III) when focusing on prospective observational data and excluding the retrospective studies.

Current evidence from meta-analyses of RCTs was limited to the beneficial effects of SUA-lowering therapy on some intermediate traits or biomarkers related to cardiovascular and renal disorders (e.g. blood pressure, endothelial functions and renal function). However, when multiple meta-analyses of RCTs existed for their traits or markers, often their results

were not concordant in direction of effect and/or statistical significance. Although 12 health outcomes had $p<0.05$, only recurrence of nephrolithiasis with citrates treatment achieved $p<10^{-3}$ with 95% PI excluding the null. Two additional health outcomes (recurrence of nephrolithiasis with thiazides treatment and end-stage renal disease in CKD patients with allopurinol treatment) also had a 95% PI excluding the null. Large heterogeneity and evidence of bias were common even in meta-analyses of RCTs (in 45.0% and 35.0% of meta-analyses of RCTs respectively). When incorporating evidence from meta-analyses of RCTs with that from meta-analyses of observational studies, there was a notable gap, as health outcomes that were investigated in meta-analyses of observational studies and classified as class I-II, have generally not been evaluated in meta-analyses of RCTs. In a few cases, data from RCTs on surrogate outcomes (e.g. SBP, DBP, renal function tests) that correspond to disease outcomes in observational studies (hypertension, CKD) were available, but conclusions from extrapolation of surrogate outcomes, which were evaluated in short-term trials, to long-term clinical outcomes should be treated with caution.

As an alternative to RCTs, MR design has been developed for exploring the causal effect of biomarkers on health outcomes. Fifty-six MR studies were identified that explored the causal role of SUA in cardiovascular, metabolic, neurocognitive and renal disorders or related traits and biomarkers. In contrast to the meta-analyses of observational studies where most of the results (76.3%) were statistically significant at $p<0.05$, the majority (83.9%) of health outcomes investigated in MR studies were not statistically significant. The generally negative results across so many health outcomes suggest that the large effects have probably not been missed, but most of the included MR studies could have been underpowered to detect modest effects. When retaining the largest MR study for each health outcome, significant results with $p<0.05$ were only reported for 9 health outcomes, and only 4 of these health outcomes (diabetic macrovascular disease, arterial stiffness [internal diameter of carotid artery], renal events, and gout) had $p<0.01$, while only the gout outcome was based on evidence from MR study with adequate power. Of the other 5 health outcomes with $p<0.05$, Parkinson's disease and memory performance had at least one other MR study that was not significant or had an association in the opposite direction.

There are a number of instrumental variable assumptions that need to be fulfilled for the results of an MR analysis to be valid. The first assumption states that the genetic instrument should be strongly associated with the intermediate phenotype. SUA has a significant heritable component with an overall heritability of 40%-60% (364), but the strength of genetic instruments used in MR studies was small or moderate, accounting for only 1.8%-6.0%

of SUA variance. Currently, the proportion of SUA variance explained by all common genetic variants identified by GWAS remains relatively small (7.0%) (151). This limits the power of genetic instruments to detect causal associations with SUA. The second and third assumptions (the instrument is associated with the outcome through the studied exposure only and the genotype is independent of other factors which affect the outcome) are more difficult to evaluate given the largely unknown complexity and interconnectedness of biologic pathways underlying the genetic variants related to SUA. The included MR studies tried to validate these assumptions either by excluding SNPs related to other known confounding factors, by excluding SNPs that had potential pleiotropic effects or by applying novel MR methods to account for pleiotropic effects (e.g. Egger or network MR).

3.5.2 Clinical implications and future research

Current recommendations on the pharmacological treatment of hyperuricaemia are related to gout or nephrolithiasis (124). Since a wide range of health outcomes have been identified to be associated with SUA, a renewed interest in whether individuals with asymptomatic hyperuricaemia should be treated with SUA-lowering drugs for the prevention or treatment of these non-crystal deposition diseases developed. In this study, we raised large uncertainty about the potential therapeutic benefits of an expansion of SUA-lowering therapy. Although we identified some highly suggestive associations from observational studies, there was a lack of concordance with clinically relevant endpoints from RCTs or surrogate endpoints from MR studies and therefore there is insufficient evidence to support any SUA-lowering drug intervention for these outcomes. Furthermore, the adverse effects of SUA-lowering drugs should be taken into consideration (for example an estimated 0.1% of patients treated with allopurinol, the first line SUA-lowering drug, develop allopurinol hypersensitivity syndrome, which can be life-threatening) (56).

On the other hand, our study does not support one of the recommendations in the recently updated EULAR gout treatment guidelines which suggests that SUA level <178 μmol/L is not recommended for gout management in the long term (55). This recommendation is based on a number of observational studies in which low SUA levels were associated with increased risk of multiple neurological diseases, including Alzheimer's and Parkinson's disease (365-367). However, in our umbrella review a number of meta-analyses reported nominally statistically significant associations of low SUA levels with increased risk of multiple neurological diseases, but several other (9 out of 28) meta-analyses did not support these findings. Moreover, our credibility assessment showed that the nominally significant

associations were consistent with Class IV evidence, and a causal effect has not consistently been established for any neurological disease in MR studies. Therefore, there is no adequate evidence against lowering SUA in gout patients in relation to an increased risk of neurological diseases.

In relation to future research, efforts to address the limitations and caveats in current evidence will be beneficial. In particular, as the current clinical trials of SUA-lowering treatment largely focus on the effect of allopurinol on some intermediate traits or biomarkers, the effect of SUA reduction on clinically relevant endpoints of the convincing and highly suggestive associations might be worth of further investigation. In addition, efforts to evaluate whether other SUA-lowering agents have the same effect as xanthine oxidase inhibitors, will help to determine if these effects are truly due to the SUA reduction *per se* rather than the mechanisms of xanthine oxidase inhibition. Finally, noting the largely discordant evidence in MR studies, better designed MR studies with collaboration of large international consortia may assist in deciding whether the lack of replication of highly suggestive findings of observational studies is due to low power to detect moderate/small effects or due to truly negative effects.

3.5.3 Strengths and weaknesses of this review

The strengths of the umbrella reviews have been previously described in detail (255, 362, 363). Here, we summarised and presented the evidence of the associations between SUA and a wide spectrum of health related outcomes systematically and thoroughly by incorporating information from meta-analyses of observational studies, meta-analyses of RCTs and MR studies. We then calculated a number of additional metrics and applied well-defined criteria to assess the credibility of the observed associations.

In relation to study weaknesses, umbrella reviews focus on existing meta-analyses and therefore outcomes that were not assessed in a meta-analysis are not included in the review. For example, we found no formal meta-analysis of observational studies on SUA and urolithiasis or gout, even though these associations are very well established. Although there are some differences in SUA levels between men and women, there is not sufficient evidence at a meta-analysis level and therefore we did not attempt to perform subgroup analyses by gender. To avoid subjectivity, we did not include reviews without explicit systematic literature searches, but this could limit the breadth of the results to some extent, if some non-systematic reviews cover questions that have not been addressed by systematic reviews (368, 369). Furthermore, we did not appraise the quality of the individual studies, since this should

be the responsibility of the authors of the original meta-analysis and it was beyond the scope of the current umbrella review.

We adopted credibility assessment criteria, which were based on established tools for observational evidence and their individual limitations have been summarised previously (255, 362, 363). None of the components of these criteria provide firm proof of lack of reliability, but they cumulatively map the possibility that the results are susceptible to bias and uncertainty. Given the wide variety of study designs and populations considered in several of the meta-analyses, one may claim that large heterogeneity in particular may not necessarily be worrisome. However, considering it is difficult to differentiate the real heterogeneity from the heterogeneity that reflects some forms of bias or uncertainty, we applied I^2<50% as one of the criteria for Class I evidence (convincing) for meta-analyses of observational studies, so as to assign the top evidence grade only to associations that are most robust and without hints of bias. In most cases I^2>50% indicates the presence of component studies with opposite effects or of component studies with and without statistically significant associations. However, nine meta-analyses of observational studies classified as class II, III or IV had an I^2>50% with all component studies reporting a statistically significant association of the same direction. Only one of these nine meta-analyses (heart failure incidence) would be upgraded from Class II to Class I, if we did not consider the heterogeneity criterion, since the other 8 also failed additional class I criteria. No meta-analyses of RCTs had an I^2>50% with all component studies reporting a statistically significant association with the same direction.

Finally, another limitation of the umbrella review approach is the use of existing meta-analyses taking their results at face value. Meta-analyses are known to have common flaws (370) and their results may also depend on choices made about what estimates to select from each primary study and how to represent them in the meta-analysis, e.g. in what contrast of exposure levels. This may be a common problem when the factor of interest is continuous, as in the case of SUA and where very different comparisons of levels of the risk factor may be selected to express risk (371). We therefore decided to investigate any meta-analyses with seemingly convincing evidence in more detail. In this process, the only meta-analysis that seemed to achieve convincing evidence (class I: stroke mortality) was found to actually have major flaws. Re-calculation of the results showed that the evidence was actually downgraded to be non-significant. It is possible that similar in-depth evaluations might have downgraded the credibility of some additional meta-analyses. In addition, we noted that many primary studies are represented in the calculations of meta-analyses by using only a small subset of

the data of extreme groups, e.g. as the risk ratio for an event in extreme quintiles of SUA. In these cases, the number of events pertinent to these extreme groups may be much fewer than the total number of events used in calculating the amount of evidence criterion. Therefore, some meta-analyses that seemingly include studies with a total of over 1000 cases may actually capture fewer than 1000 cases in the main calculations and thus their grading appraisal should have been weaker. These flaws and deficiencies are very difficult to decipher without an in-depth re-construction from scratch of all observational meta-analyses and they may explain why observational evidence for SUA associations generally did not show good concordance with randomised trial and MR evidence in our umbrella evaluation.

Meta-analyses of observational data for SUA and other risk factors need to be strengthened. For continuous putative risk factors such as SUA, a wide consensus on the categorisation of levels of interest would be useful to achieve and careful meta-analyses of individual level data in inclusive consortia may help achieve this. This approach would allow a more accurate and reliable exploration of both linear and non-linear associations, e.g. the possibility of U-shaped associations with increased risk at both very high and very low levels. Currently available data from meta-analyses do not allow for consistent handling and assessment of such non-linear relationships. Conversely, data dredging using different categorisations of levels and exposure contrasts is likely to be fuelling a literature with spurious associations (372).

3.6 Conclusion

This comprehensive umbrella review will help investigators to judge the relative priority of health outcomes related to SUA in relation to future research and to clinical management of disease. In summary, despite a few hundred systematic reviews, meta-analyses and MR studies exploring 136 unique health outcomes, convincing evidence of a clear role of SUA exists for only gout and nephrolithiasis. Concordant evidence between observational studies and RCTs existed for hypertension and CKD, but a potential causal role of SUA for these outcomes has not been verified by current MR studies and even for these two outcomes not all meta-analyses of RCTs are concordant among themselves and with observational evidence. Therefore, the available evidence does not support any change in the existing clinical recommendations in relation to hyperuricaemia.

3.7 Supplementary information

Supplementary Table 3 - 1: Keywords and search strategy used in the umbrella review.

MEDLINE (OvidSP)
1. Uric acid.mp. or Uric Acid/
2. Uric Acid/ or urate*.mp.
3. Acid uric.mp. or Uric Acid/
4. Hyperuricaemia.mp. or Uric Acid/ or Hyperuricaemia/
5. Uric Acid/ or hypouricaemia.mp.
6. Uric Acid/ or hyperuricosuria.mp.
7. Uric Acid/ or hypouricosuria.mp.
8. 1 or 2 or 3 or 4 or 5 or 6 or 7
9. "Review Literature as Topic"/ or "Review"/ or review*.mp.
10. Meta-Analysis as Topic/ or meta-analys*.mp.
11. Mendelian Randomisation Analysis/ or Mendelian randomi*.mp.
12. 9 or 10 or 11
13. 8 and 12
EMBASE (OvidSP)
1. Uric acid.mp. or uric acid/
2. Urate*.mp. or urate/
3. Uric acid/ or acid uric.mp.
4. Hyperuricaemia.mp. or hyperuricaemia/
5. Hypouricaemia.mp. or hypouricaemia/
6. Hyperuricosuria.mp. or hyperuricosuria/

7. Uric acid/ or hypouricosuria.mp. 8. 1 or 2 or 3 or 4 or 5 or 6 or 7 9. Systematic review.mp. or "systematic review"/ 10. "Systematic review"/ or meta-analys*.mp. 11. Mendelian randomisation analysis/ or Mendelian randomi*.mp. or Mendelian randomisation/ 12. 9 or 10 or 11 13. 8 and 12
Cochrane library (uric acid OR acid uric OR urate* OR hyperuricaemia OR hypouricaemia OR hyperuricosuria OR hypouricosuria) AND (systematic review* OR meta-analys*)

Supplementary Table 3 - 2: General characteristics and main findings of the systematic reviews of observational studies.

Author	Year	Population	Study design	Comparison	Outcome	N studies	Authors' interpretation
Cardiovascular Outcomes							
Baker(373)	2005	Healthy subjects	Cohort	SUA level	CVD	10	The excess risk associated with SUA in healthy patients is likely to be small. High SUA is likely an independent risk factor in patients at high CVD risk.
Baker(373)	2005	Patients at high CVD risk	Cohort	SUA level	CVD	11	
Strazzullo(374)	2007	General	Cohort	SUA level	CVD	16	SUA is a very weak predictor of CVD in healthy population, but a significant independent predictor among subjects at high or very high risk.
Strazzullo(374)	2007	Patients with arterial hypertension	Cohort	SUA level	CVD	8	
Strazzullo(374)	2007	Patients at high CVD risk	Cohort	SUA level	CVD	5	
Barron(260)	2015	General	Cohort	SUA level	CVD	3	Greater risk of CVD/all-cause mortality in those with the highest than with the lowest quartiles of SUA.
Dimitroula(261)	2008	General	Cohort	SUA level	Stroke	13	It remains controversial whether elevated serum uric acid is neuroprotective or injurious at the onset of acute stroke.
Hwu(262)	2010	General	Cross-sectional	SUA level	Hypertension	6	All except one study have documented a direct association with either incident hypertension or increase in blood pressure.
Hwu(262)	2010	General	Cohort/case-control	Hyper vs normal	Hypertension	17	
Kidney disease							
Avram(263)	2008	General	Cohort	SUA level	Kidney dysfunction	9	Nearly all published prospective studies support the role of hyperuricaemia as an independent risk factor for renal dysfunction.
Feig(375)	2009	General	Observational	SUA level	Chronic kidney disease (CKD)	11	The preponderance of epidemiological evidence suggests a direct link between uric acid and CKD.
Cognitive diseases							
Alonso(376)	2010	General	Cohort/case-control	SUA level	Parkinson's disease (PD)	10	PD patients have lower SUA levels than controls; SUA is strongly and linearly associated with reduced risk of PD.
Alonso(376)	2010	PD patients	Cohort/case-control	SUA level	Clinical disability and	3	

					worse cognitive performance		
Chang(377)	2014	AD patients vs control	Case-control	SUA level	Alzheimer's disease (AD)	6	Three studies elaborate that plasma or serum uric acid level is significantly lower in AD, while three other studies do not observe this difference.
Other Outcomes							
Cnossen(378)	2006	Women	Cohort	SUA level	Pre-eclampsia	5	There is currently insufficient evidence to draw firm conclusions about the accuracy of serum uric acid determination in predicting pre-eclampsia.

Supplementary Table 3 - 3: General characteristics and main findings of the 144 meta-analyses of observational studies.

Author	Year	Outcome	Population	Comparison	N Studies	N Participants	N Cases	Type of metric	Effect model reported	Reported summary effect (95% CI)	Concordance in overlapped MA
Cardiovascular outcomes											
Tamariz(310)	2014	Atrial fibrillation incidence	General	Highest vs lowest SUA category	3	138,306	3,466	RR	Random	1.67 (1.23, 2.27)	**YES(268, 310, 379)**
Xu(379)	2015	Atrial fibrillation incidence	General	Hyper vs normal	7	85,530	NA	RR	Random	1.80 (1.37, 2.38)	
Xu(379)	2015	Atrial fibrillation new-onset	General	Hyper vs normal	5	84,837	NA	RR	Random	1.66 (1.22, 2.26)	
Zhang(268)	2016	Atrial fibrillation incidence	General	Hyper vs normal	6	426,159	7,595	RR	Random	1.49 (1.24, 1.79)	
Tamariz(310)	2014	Atrial fibrillation	Atrial fibrillation vs controls	SUA level (mg/dL)	6	7,930	1,603	SMD	Random	0.42 (0.27, 0.58)	
Zhao(269)	2016	Atrial fibrillation recurrence	Atrial fibrillation patients	Hyper vs normal	4	1,298	393	OR	Random	1.37 (0.98, 1.93)	
Xu(379)	2015	Atrial fibrillation recurrence	General	Hyper vs normal	2	61,955	NA	aRR	Random	2.07 (1.61, 2.67)	
Wheeler(380)	2005	CHD incidence	General	Highest vs lowest SUA category	16	164,542	9,485	RR	Random	1.13 (1.07, 1.20)	**YES(16, 380-382)**
Kim(381)	2010	CHD	General	Hyper vs normal	9	53,750	5,113	aRR	Random	1.09 (1.03, 1.16)	
Braga(382)	2015	CHD incidence	General	Hyper vs normal	9	457,915	19,119	RR	Random	1.21 (1.07, 1.36)	
Li(16)	2016	CHD incidence	General	Hyper vs normal	13	70,382	6,666	aRR	Random	1.13 (1.05, 1.21)	
Qin(270)	2016	CVD	Hypertensive patients	Hyper vs normal	6	19,546	1,054	aHR	Random	1.17 (1.07, 1.27)	
Qin(270)	2016	CVD	Hypertensive patients	Continuous SUA	4	NA	NA	aHR	Random	1.51 (1.13, 2.03)	
Huang(383)	2014	Heart failure incidence	General	Hyper vs normal	5	427,917	101,71	HR	Random	1.65 (1.41, 1.94)	
Huang(383)	2014	Heart failure incidence	General	Highest vs lowest SUA category	4	NA	NA	HR	Random	1.64 (1.39, 1.94)	

Author	Year	Outcome	Population	Comparison	N Studies	N Participants	N Cases	Type of metric	Effect model reported	Reported summary effect (95% CI)	Concordance in overlapped MA
Huang(383)	2014	Heart failure incidence	General	1 mg/dL SUA increase	3	420,756	9,812	HR	Fixed	1.19 (1.17, 1.21)	
Zhang(384)	2009	Hypertension	General	Highest vs lowest SUA category	8	28,657	NA	RR	Random	1.55 (1.32, 1.82)	**YES(272, 384, 385)**
Grayson(385)	2011	Hypertension incidence	General	Hyper vs normal	12	32,390	NA	RR	Random	1.81 (1.55, 2.07)	
Grayson(385)	2011	Hypertension incidence	General	Hyper vs normal	11	33,925	NA	aRR	Random	1.41 (1.23, 1.58)	
Grayson(385)	2011	Hypertension incidence	General	1 mg/dL SUA increase	6	23,018	NA	aRR	Random	1.13 (1.06, 1.20)	
Grayson(385)	2011	Hypertension incidence	General	1 SD SUA increase	8	30,492	NA	RR	Random	1.16 (1.07, 1.26)	
Wang(272)	2014	Hypertension incidence	General	Hyper vs normal	17	71,630	18,751	aRR	Random	1.48 (1.33, 1.65)	
Wang(272)	2014	Hypertension incidence	General	1 SD SUA increase	10	37,125	7,584	aRR	Random	1.19 (1.11, 1.28)	
Wang(272)	2014	Hypertension incidence	General	1 mg/dL SUA increase	5	15,951	4,941	aRR	Random	1.15 (1.06, 1.26)	
Jiang(273)	2016	Prehypertension	General	Highest vs lowest SUA category	8	44,095	20,832	OR	Random	1.84 (1.42, 2.38)	
Zhang(268)	2016	Left atrial thrombus or spontaneous echo contrast (LATH/LASEC)	Patients with mitral stenosis, sinus rhythm or atrial fibrillation	Highest vs lowest SUA category	6	2,381	241	OR	Random	1.59 (1.13, 2.23)	
Yan(386)	2014	Major adverse cardiovascular events (MACE)	Acute myocardial infarction (AMI) patients vs controls	Hyper vs normal	6	2,406	651	RR	Fixed	3.44 (2.33, 5.08)	
Song(387)	2015	MACE	Patients after PCI	Hyper vs normal	2	3,054	NA	RR	Fixed	1.78 (1.26, 2.52)	
Trkulja(388)	2012	Medium term MACE	Patients with acute myocardial infarction	Highest vs lowest SUA category	4	4,299	1,240	OR	Random	1.62 (1.20, 2.19)	

Author	Year	Outcome	Population	Comparison	N Studies	N Participants	N Cases	Type of metric	Effect model reported	Reported summary effect (95% CI)	Concordance in overlapped MA
Trkulja(388)	2012	Short term MACE	Patients with acute myocardial infarction	Highest vs lowest SUA category	7	6,470	787	OR	Random	2.21 (1.64, 2.97)	
Qin(270)	2016	Stroke	Hypertensive patients	Hyper vs normal	5	NA	NA	HR	Random	0.94 (0.67, 1.33)	
Qin(270)	2016	Stroke	Hypertensive patients	Continuous SUA level	3	9,978	217	aHR	Random	1.11 (0.98, 1.16)	
Kim(389)	2009	Stroke incidence	General	Hyper vs normal	6	11,495	NA	RR	Random	1.41 (1.05, 1.76)	**YES(276, 389)**
Li(276)	2014	Stroke incidence	General	Hyper vs normal	5	24,548	1,290	aRR	Random	1.22 (1.02, 1.46)	
Diabetes related outcomes											
Jia(390)	2013	Impaired fasting glucose or T2DM	General	Highest vs lowest SUA category	12	62,834	6,340	RR	Random	1.57 (1.39, 1.77)	
Kodama(277)	2009	Type 2 diabetes (T2DM)	General	1 mg/dL SUA increase	11	42,834	3,305	RR	Random	1.17 (1.09, 1.25)	**YES(277, 307, 390)**
Jia(390)	2013	T2DM	General	Highest vs lowest SUA category	9	48,808	5,115	RR	Fixed	1.67 (1.51, 1.86)	
Lv(307)	2013	T2DM incidence	General without diabetes	Hyper vs normal	8	32,016	2,930	RR	Fixed	1.56 (1.39, 1.76)	
Lv(307)	2013	T2DM incidence	General without diabetes	1 mg/dL SUA increase	6	21,592	2,203	RR	Fixed	1.06 (1.04, 1.07)	
Qin(270)	2016	Diabetes incidence	Hypertensive patients	Hyper vs normal	2	8,247	564	aHR	Random	1.84 (1.02, 3.30)	
Qin(270)	2016	Diabetes incidence	Hypertensive patients	Continuous SUA level	2	NA	NA	aHR	Random	1.28 (1.18, 1.38)	
Xu(279)	2013	Diabetic nephropathy	T2DM patients	Continuous/categorical SUA level	3	3,166	196	OR	Random	1.91 (1.07, 3.42)	
Xu(279)	2013	Diabetic retinopathy	T2DM patients	Continuous/categorical SUA level	2	1,739	311	OR	Random	1.23 (0.81, 1.87)	
Xu(279)	2013	Diabetic vascular complications	T2DM patients	Continuous/categorical SUA level	6	5,017	967	OR	Random	1.28 (1.12, 1.46)	

Author	Year	Outcome	Population	Comparison	N Studies	N Participants	N Cases	Type of metric	Effect model reported	Reported summary effect (95% CI)	Concordance in overlapped MA
Xu(279)	2013	Diabetic microvascular complications	T2DM patients	Continuous/categorical SUA level	5	4,513	854	OR	Random	1.47 (1.11, 1.94)	
Xu(279)	2013	Peripheral vascular disease	T2DM patients	Continuous/categorical SUA level	3	2,538	151	OR	Random	1.27 (0.91, 1.78)	
Xu(279)	2013	Diabetic Macrovascular disease	T2DM patients	Continuous/categorical SUA level	3	2,538	187	OR	Fixed	1.03 (1.00, 1.06)	
Xu(279)	2013	Diabetic neuropathy	T2DM patients	Continuous/categorical SUA level	2	2,034	231	OR	Random	1.19 (0.61, 2.32)	**NO(279, 281)** (discordance in statistical significance)
Yu(281)	2016	Diabetic peripheral neuropathy	Diabetic peripheral peuropathy patients vs diabetic controls	SUA level (μmol/L)	7	2,035	464	MD	Random	50.03 (22.14, 77.93)	
Yu(281)	2016	Diabetic peripheral neuropathy	Diabetic peripheral peuropathy patients vs diabetic controls	Hyper vs normal	5	4,097	894	RR	Random	2.83 (2.13, 3.76)	
Kidney disorders											
Li(305)	2011	Impaired kidney function	General	Hyper vs normal	3	3,004	NA	RR	Random	1.35 (1.12, 1.63)	
Li(305)	2011	CKD incidence	General	Hyper vs normal	10	276,801	3,730	aRR	Random	1.49 (1.27, 1.75)	**YES(282, 283, 305)**
Zhu(282)	2014	CKD incidence	Middle-aged Populations	1 mg/dL SUA increase	15	99,205	3,492	RR	Random	1.22 (1.16, 1.28)	
Li(283)	2014	CKD new-onset incidence	Non-CKD population	Hyper vs normal	6	28,256	NA	HR	Fixed	2.59 (2.14, 3.13)	
Li(283)	2014	CKD new-onset incidence	Non-CKD population	1 mg/dL SUA increase	7	153,620	7,014	HR	Random	1.06 (1.04, 1.08)	
Li(283)	2014	CKD new-onset incidence	Healthy population	Hyper vs normal	4	NA	NA	HR	Random	2.86 (2.30, 3.56)	
Li(283)	2014	CKD new-onset incidence	Diabetic patients	Hyper vs normal	2	NA	NA	HR	Random	1.90 (1.30, 2.78)	
Huang(284)	2012	eGFR (mL/min·1.73 m^2)	Renal transplant recipients	Hyper vs normal	8	2,075	NA	MD	Random	-11.24 (-16.34, -6.14)	

Author	Year	Outcome	Population	Comparison	N Studies	N Participants	N Cases	Type of metric	Effect model reported	Reported summary effect (95% CI)	Concordance in overlapped MA
Huang(284)	2012	SCr (μmol/L)	Renal transplant recipients	Hyper vs normal	5	873	NA	MD	Fixed	0.24 (0.17, 0.31)	
Huang(284)	2012	Graft loss	Renal transplant recipients	Hyper vs normal	3	910	154	OR	Fixed	2.29 (1.55, 3.39)	
Huang(284)	2012	Graft loss	Renal transplant recipients	1 mg/dL SUA increase	3	1,050	NA	aHR	Fixed	1.21 (1.08, 1.37)	
Huang(284)	2012	Chronic allograft nephropathy	Renal transplant recipients	Hyper vs normal	4	1,057	113	OR	Fixed	2.85 (1.85, 4.38)	
Neurocognitive disorders											
Khan(391)	2013	Alzheimer's disease (AD)	AD patients vs controls	SUA level (mg/dL)	17	3,447	1,153	SMD	Random	-0.42 (-0.62, -0.21)	**NO(301, 391-393)** (discordance in statistical significance)
Schrag(392)	2013	Alzheimer's disease (AD)	AD patients vs controls	SUA level (mg/dL)	10	900	453	Hedge's G	Random	-0.59 (-1.26, -0.09)	
Chen(301)	2014	Alzheimer's disease (AD)	AD patients vs controls	SUA level (mg/dL)	11	2,708	647	SMD	Random	-0.50 (-1.23, 0.22)	
Du(393)	2016	Alzheimer's disease (AD)	AD patients vs controls	SUA level (mg/dL)	21	3,617	1,128	MD	Random	-0.77 (-1.18, -0.36)	
Du(393)	2016	Alzheimer's disease (AD)	Alzheimer's patients vs controls	Highest vs lowest SUA category	3	7,372	NA	RR	Fixed	0.66 (0.52, 0.85)	
Bartoli(394)	2016	Bipolar disorder	Bipolar disorder vs controls	SUA level (mg/dL)	9	1,127	619	SMD	Random	0.65 (0.33, 0.97)	
Bartoli(394)	2016	Bipolar disorder	Bipolar disorder vs major depression	SUA level (mg/dL)	5	735	399	SMD	Random	0.46 (0.16, 0.75)	
Bartoli(394)	2016	Bipolar disorder phase (Depressive vs euthymic phase)	Depressive vs euthymic phase	SUA level (mg/dL)	6	375	NA	SMD	Random	-0.11 (-0.33, 0.11)	
Bartoli(394)	2016	Bipolar disorder phase	Manic/mixed vs depressive phase	SUA level (mg/dL)	7	472	NA	SMD	Random	0.34 (0.02, 0.66)	
Bartoli(394)	2016	Bipolar disorder phase	Manic/mixed vs euthymic phase	SUA level (mg/dL)	6	402	NA	SMD	Random	0.19 (-0.10, 0.49)	
Khan(391)	2013	Dementia/cognitive impairment	Dementia/cognitive impairment patients vs controls	SUA level (mg/dL)	31	7,021	2,681	SMD	Random	-0.33 (-0.48, -0.17)	

Author	Year	Outcome	Population	Comparison	N Studies	N Participants	N Cases	Type of metric	Effect model reported	Reported summary effect (95% CI)	Concordance in overlapped MA
Khan(391)	2013	Dementia/cognitive impairment	Dementia/cognitive impairment patients vs controls	Highest vs lowest SUA category	5	3,281	489	OR	Random	1.18 (0.96, 1.46)	
Khan(391)	2013	Vascular dementia (VaD)	Vascular dementia patients vs controls	SUA level (mg/dL)	7	597	272	SMD	Random	-0.05 (-0.88, 0.78)	
Schrag(392)	2013	Mild cognitive impairment (MCI)	MCI patients vs controls	SUA level (mg/dL)	2	129	49	Hedge's G	Random	-0.73 (-2.7, 1.24)	**YES(391, 392)**
Khan(391)	2013	Mild cognitive impairment (MCI)	MCI patients vs controls	SUA level (mg/dL)	4	731	515	SMD	Random	-0.24 (-0.90, 0.42)	
Khan(391)	2013	Mixed or undifferentiated dementia	Mixed or undifferentiated dementia vs controls	SUA level (mg/dL)	4	1,998	NA	SMD	Random	0.19 (-0.17, 0.54)	
Khan(391)	2013	Parkinson's disease (PD)	PD patients vs controls	SUA level (mg/dL)	7	1,008	514	SMD	Random	-0.67 (-1.05, -0.29)	**YES(309, 391, 395)**
Shen(395)	2013	Parkinson's disease (PD)	PD patients vs controls	SUA level (μM/L)	6	2,493	1,217	SMD	NA	-0.52 (-0.72, -0.31)	
Shen(395)	2013	Parkinson's disease (PD)	General	Hyper vs normal	6	33,185	578	RR	Random	0.65 (0.43, 0.97)	
Shen(395)	2013	Parkinson's disease (PD)	General	Hyper vs normal	3	11,795	NA	RR(dose-respond)	NA	0.77 (0.68, 0.88)	
Shen(309)	2013	PD progression	PD patients	Hyper vs normal	2	1,578	NA	RR	Fixed	0.56 (0.43, 0.72)	
Liu(396)	2012	Multiple sclerosis (MS)	MS patients vs controls	SUA level (umol/L)	8	1,037	556	SMD	Random	-0.68 (-0.82, -0.55)	**YES(396, 397)**
Wang(397)	2016	Multiple sclerosis (MS)	MS patients vs control	SUA level (μmol/L)	10	2,216	1,308	SMD	Random	-0.40 (-0.73, -0.07)	
Wang(397)	2016	Neruomyelistsopticis (NMO)	NMO patients vs control	SUA level (μmol/L)	3	1,137	229	SMD	Random	-0.85 (-1.24, -0.46)	
Wang(397)	2016	MS and NMO	MS+NMO patients vs control	SUA level (μmol/L)	13	2,445	1,537	SMD	Random	-0.52 (-0.81, -0.24)	
Abraham(398)	2014	Amyotrophic lateral sclerosis (ALS)	ALS patients vs controls	SUA level (mg/dL)	3	826	311	Hedge's G	NA	0.84 (NA, NA)	

Author	Year	Outcome	Population	Comparison	N Studies	N Participants	N Cases	Type of metric	Effect model reported	Reported summary effect (95% CI)	Concordance in overlapped MA
Flatow(399)	2013	Schizophrenia (Chronic)	Chronic Schizophrenia patients vs controls	SUA level (mg/dL)	3	241	103	Hedge's G	Random	p=0.15	
Flatow(399)	2013	Schizophrenia (First-Episode Psychosis)	Schizophrenia patients in First-Episode Psychosis vs controls	SUA level (mg/dL)	2	274	155	Hedge's G	Random	p<0.01	
Cancer outcomes											
Yan(400)	2015	Cancer incidence	General	Highest vs lowest SUA category	5	456,053	14,355	RR	Fixed	1.03 (1.01, 1.05)	
Yan(400)	2015	Cancer in Digestive organs incidence	General	Highest vs lowest SUA category	3	266,347	2,521	RR	Random	1.08 (0.94, 1.25)	
Yan(400)	2015	Cancer in Lymphoid and hematopoietic systems incidence	General	Highest vs lowest SUA category	2	86,739	397	RR	Fixed	1.71 (1.10, 2.68)	
Yan(400)	2015	Cancer in Male genital organs incidence	General	Highest vs lowest SUA category	3	162,022	2,634	RR	Fixed	1.06 (1.00, 1.13)	
Yan(400)	2015	Cancer in Respiratory system and intrathoracic organs incidence	General	Highest vs lowest SUA category	4	456,053	2,941	RR	Random	1.05 (0.93, 1.19)	
Yan(400)	2015	Cancer in Urinary organs incidence	General	Highest vs lowest SUA category	2	86,739	536	RR	Random	1.17 (0.44, 3.15)	
All-cause and cause-specific mortality											
Kim(381)	2010	CHD mortality	General	Hyper vs normal	8	253,336	4,473	aRR	Random	1.16 (1.01, 1.30)	**YES(16, 381, 382)**
Kim(381)	2010	CHD mortality	General	1 mg/dL SUA increase	4	102,342	770	aRR	Random	1.12 (1.05, 1.19)	
Braga(382)	2015	CHD mortality	General	Hyper vs normal	6	237,421	5,572	RR	Random	1.21 (1.00, 1.46)	
Li(16)	2016	CHD Mortality	General	Hyper vs normal	13	876,584	24,198	aRR	Random	1.27 (1.16, 1.39)	

Author	Year	Outcome	Population	Comparison	N Studies	N Participants	N Cases	Type of metric	Effect model reported	Reported summary effect (95% CI)	
Li(16)	2016	CHD Mortality	General	1 mg/dL SUA increase	6	NA	NA	RR	Random	1.15 (1.09, 1.21)	
Qin(270)	2016	CVD mortality	Hypertensive patients	Hyper vs normal	3	NA	NA	aHR	Random	1.31 (0.96, 1.78)	
Huang(383)	2014	CVD mortality	Heart failure patients	Hyper vs normal	2	2,250	NA	HR	Random	1.45 (1.18, 1.78)	
Zhao(293)	2014	CVD mortality	General	Highest vs lowest SUA category	9	165,806	6,121	RR	Random	1.37 (1.19, 1.57)	**YES(293, 401)**
Yang(401)	2015	CVD mortality	General	Highest vs lowest SUA category	3	105,329	1,829	RR	Random	1.25 (1.00, 1.56)	
Kim(389)	2009	Stroke mortality	General	Hyper vs normal	6	45,751	NA	aRR	Random	1.26 (1.12, 1.39)	**YES(276, 389)**
Li(276)*	2014	Stroke mortality	General	Hyper vs normal	9	1,017,810	21,281	aRR	Random	1.33 (1.24, 1.43)	
Qin(270)	2016	Stroke mortality	Hypertensive patients	Continuous SUA level	2	NA	NA	aHR	Random	1.20 (0.95, 1.51)	
Xia(294)	2016	CKD mortality	General	Highest vs lowest SUA category	14	15,930	3,245	aHR	Random	1.52 (1.33, 1.73)	
Xia(294)	2016	CKD mortality	General	1 mg/dL SUA increase	21	23,443	3,904	aHR	Random	1.08 (1.04, 1.11)	
Yan(400)	2015	Cancer mortality	General	Highest vs lowest SUA category	12	632,472	NA	RR	Random	1.17 (1.04, 1.32)	
Yan(400)	2015	Cancer mortality in bone, connective tissue, soft tissue, and skin	General	Highest vs lowest SUA category	NA	112,296	NA	RR	Fixed	0.94 (0.47, 1.87)	
Yan(400)	2015	Cancer mortality in digestive organs	General	Highest vs lowest SUA category	4	187,886	855	RR	Fixed	1.27 (1.08, 1.49)	
Yan(400)	2015	Cancer mortality in lymphoid and hematopoietic systems	General	Highest vs lowest SUA category	NA	112,296	NA	RR	Fixed	1.18 (0.82, 1.70)	
Yan(400)	2015	Cancer mortality in male genital organs	General	Highest vs lowest SUA category	NA	88,033	NA	RR	Random	0.51 (0.07, 3.85)	

Author	Year	Outcome	Population	Comparison	N Studies	N Participants	N Cases	Type of metric	Effect model reported	Reported summary effect (95% CI)	Concordance in overlapped MA
Yan(400)	2015	Cancer mortality in respiratory system and intrathoracic organs	General	Highest vs lowest SUA category	2	116,646	164	RR	Random	1.08 (0.61, 1.91)	
Yan(400)	2015	Cancer mortality in urinary organs	General	Highest vs lowest SUA category	2	112,296	NA	RR	Fixed	1.35 (0.88, 2.07)	
Tamariz(402)	2009	All-cause mortality	Heart failure patients	Hyper vs normal	6	1,456	NA	RR	Fixed	2.13 (1.78, 2.55)	**YES(383, 402)**
Tamariz(402)	2009	All-cause mortality	Acute heart failure patients	Hyper vs normal	4	772	NA	RR	Fixed	2.40 (1.50, 3.70)	
Tamariz(402)	2009	All-cause mortality	Chronic heart failure patients	Hyper vs normal	2	772	NA	RR	Fixed	2.10 (1.50, 2.90)	
Huang(383)	2014	All-cause mortality	Heart failure patients	Hyper vs normal	11	12,444	1,888	HR	Random	2.15 (1.64, 2.83)	
Huang(383)	2014	All-cause mortality	Heart failure patients	1 mg/dL SUA increase	10	21,119	5,755	HR	Random	1.04 (1.02, 1.06)	
Li(305)	2011	All-cause mortality	CKD population	Hyper vs normal	5	1,789	609	RR	Random	1.67 (1.29, 2.16)	
Trkulja(388)	2012	Short-term mortality	AMI patients	Highest vs lowest SUA category	8	6,805	396	OR	Random	2.95 (2.29, 3.80)	
Trkulja(388)	2012	Medium term mortality	AMI patients	Highest vs lowest SUA category	5	5,194	565	OR	Random	2.28 (1.82, 2.86)	
Yan(386)	2014	In-hospital mortality	AMI patients vs controls	Hyper vs normal	6	5,686	218	RR	Random	2.10 (1.03, 4.26)	
Xu(279)	2013	Mortality	T2DM patients	Hyper vs normal	3	5,534	NA	HR	Random	1.09 (1.03, 1.16)	
Zhao(293)	2014	All-cause mortality	General	Highest vs lowest SUA category	10	143,483	7,031	RR	Random	1.24 (1.09, 1.42)	**YES(293, 401)**
Yang(401)	2015	All-cause mortality	General	Highest vs lowest SUA category	6	126,702	12,863	RR	Random	1.17 (1.03, 1.32)	
Song(387)	2015	Mortality	Patients after PCI	Hyper vs normal	9	17,268	NA	RR	Fixed	1.31 (1.21, 1.42)	
Song(387)	2015	Mortality	Patients after PCI	1 mg/dL SUA increase	3	NA	NA	RR	Random	1.25 (1.13, 1.39)	

Author	Year	Outcome	Population	Comparison	N Studies	N Participants	N Cases	Type of metric	Effect model reported	Reported summary effect (95% CI)	Concordance in overlapped MA
Qin(270)	2016	All-cause mortality	Hypertensive patients	Hyper vs normal	4	46,103	5,820	aHR	Random	1.12 (1.02, 1.23)	
Qin(270)	2016	All-cause mortality	Hypertensive patients	Continuous SUA level	3	NA	NA	aHR	Random	1.05 (0.98, 1.13)	
Other outcomes											
Trkulja(388)	2012	Medium/long-term occurrence of poor outcomes (death/MACE)	AMI patients	Highest vs lowest SUA category	3	2,683	NA	HR	Random	1.30 (1.01, 1.68)	
Trkulja(388)	2012	Medium/long-term occurrence of poor outcomes (death/MACE)	AMI patients	50 μmol/L increase	4	3,533	NA	aHR	Random	1.19 (1.03, 1.37)	
Trkulja(388)	2012	Short-term occurrence of poor outcomes (death/MACE)	AMI patients	Highest vs lowest SUA category	4	3,625	336	aOR	Random	2.26 (1.85, 2.77)	
Huang(383)	2014	Combined death or cardiac events	Heart failure patients	Hyper vs normal	9	12,699	1,765	HR	Random	1.39 (1.18, 1.63)	
Huang(383)	2014	Combined death or cardiac events	Heart failure patients	1 mg/dL SUA increase	4	2,514	NA	HR	Random	1.28 (0.97, 1.70)	
Song(387)	2015	Adverse outcomes (mortality, MAGE, In-stent restenosis)	Patients after PCI	Hyper vs normal	12	21,030	NA	RR	Random	1.46 (1.29, 1.65)	
Wang(297)	2016	Occurrence of poor outcomes	Acute ischaemic stroke patients	Highest vs lowest SUA category	9	7,932	NA	HR	Random	0.77 (0.68, 0.88)	
Wang(297)	2016	Occurrence of poor outcomes	Acute ischaemic stroke patients	SUA level (μmol/L)	4	1,879	631	MD	Fixed	30.61 (20.13, 41.08)	
Li(298)	2016	Psoriasis	Psoriasis patients vs controls	SUA level (mg/dL)	13	29,037	1,644	MD	Random	0.89 (0.05, 1.73)	
Li(298)	2016	Psoriasis severity	Sever psoriasis patients vs controls	SUA level (mg/dL)	3	300	104	MD	Random	0.53 (-1.04, 2.10)	
Zhou(299)	2016	Non-alcoholic fatty liver disease (NAFLD)	General	Highest vs lowest SUA category	9	55,573	10,581	OR	Random	1.92 (1.59, 2.31)	

Abbreviations: MA, meta-analysis; CHD, coronary heart disease; CVD, cardiovascular disease; MACE, major adverse cardiovascular events; AMI, acute myocardial infarction; PCI, percutaneous coronary intervention; T2DM, Type 2 diabetes; CKD, chronic kidney disease; AD, Alzheimer's disease; PD, Parkinson's disease; VaD, Vascular dementia; MCI, mild cognitive impairment; MS, multiple sclerosis; NMO, neruomyelistsopticis; ALS, amyotrophic lateral sclerosis; NAFLD, non-alcoholic fatty liver disease; MD, mean difference; SMD, standard mean difference; NA, not available.
* We corrected the errors and inappropriateness of the original meta-analysis, when conducted the quantitative analysis.

Supplementary Table 3 - 4: General characteristics and main findings of the 31 meta-analyses of RCTs.

Author	Year	Population	SUA-lowering therapy	Versus	Outcome	N studies	N participants	N Cases	Type of metric	Effect model	Reported summary effect (95%)	Concordance in overlapped MA
Kidney diseases												
Fink(403)	2013	Nephrolithiasis patients	Thiazides	Placebo/no treatment	Nephrolithiasis recurrence	5	300	151	RR	Random	0.52 (0.39, 0.69)	
Fink(403)	2013	Nephrolithiasis patients	Citrates	Placebo/no treatment	Nephrolithiasis recurrence	4	197	90	RR	Random	0.25 (0.14, 0.44)	
Fink(403)	2013	Nephrolithiasis patients	Allopurinol	Placebo/no treatment	Nephrolithiasis recurrence	2	152	78	RR	Random	0.59 (0.42, 0.84)	
Wang(404)	2013	General	All active therapy	Placebo/no treatment	SCr (mg/dL)	9	580	NA	SMD	Random	-1.25 (-1.98, -0.52)	
Wang(404)	2013	General	All active therapy	Placebo/no treatment	eGFR ($ml/min/1.73m^2$)	3	218	NA	SMD	Fixed	0.41 (0.14, 0.68)	
Zhang(405)	2014	CKD Patients	Allopurinol	Placebo/no treatment	SCr (μmol/L)	6	354	177	MD	Random	-62.55 (-98.10, -26.99)	YES(405, 406)
Bose(406)	2014	Patients with CKD or decreased kidney function	Allopurinol	Placebo/no treatment	SCr (mg/dL)	3	130	NA	MD	Random	-0.40 (-0.80, 0.00)	
Zhang(405)	2014	CKD Patients	Allopurinol	Placebo/no treatment	eGFR ($ml/min/1.73m^2$)	2	184	96	MD	Fixed	5.65 (1.88, 9.41)	NO(405, 406) (discordance in statistical significance)
Bose(406)	2014	Patients with CKD or decreased kidney function	Allopurinol	Placebo/no treatment	eGFR ($ml/min/1.73m^2$)	5	346	NA	MD	Random	3.10 (-0.90, 7.10)	
Zhang(405)	2014	CKD Patients	Allopurinol	Placebo/no treatment	End-stage renal disease	5	267	132	RR	Fixed	0.30 (0.19, 0.46)	NO(405, 406) (discordance in direction and statistical significance)
Bose(406)	2014	Patients with CKD or decreased kidney function	Allopurinol	Placebo/no treatment	End-stage kidney disease	2	164	NA	RR	Random	1.01 (0.15, 6.98)	
Zhang(405)	2014	CKD Patients	Allopurinol	Placebo/no treatment	Blood urea nitrogen (mmol/L)	3	169	83	MD	Fixed	-6.15 (-8.17, -4.13)	
Zhang(405)	2014	CKD Patients	Allopurinol	Placebo/no treatment	24-h urinary protein (g/day)	3	184	94	MD	Fixed	0.13 (0.28, 0.02)	

Author	Year	Population	SUA-lowering therapy	Versus	Outcome	N studies	N participants	N Cases	Type of metric	Effect model	Reported summary effect (95%)	Concordance in overlapped MA
Zhang(405)	2014	CKD Patients	Allopurinol	Placebo/no treatment	Stable renal function	5	267	132	RR	Fixed	1.73 (1.44, 2.09)	
Bose(406)	2014	Patients with CKD or decreased kidney function	Allopurinol	Placebo/no treatment	Proteinuria (g/day)	5	250	NA	MD	Random	-0.20 (-0.60, 0.10)	
Endothelial function												
Higgins(407)	2010	Population with vascular diseases	Xanthine Oxidase	Placebo/no treatment	Flow-mediated dilatation (%)	5	144	75	MD	Random	2.50 (0.15, 4.84)	
Higgins(407)	2010	Population with vascular diseases	Xanthine Oxidase	Placebo/no treatment	Forearm blood flow	5	148	74	MD	Random	68.80 (18.70, 118.90)	
Kanbay(408)*	2014	General	Allopurinol	Placebo/no treatment	Flow-mediated dilatation (%)	6	285	142	MD	Fixed	2.75 (2.49, 3.01)	
Kanbay(408)*	2014	General	Allopurinol	Placebo/no treatment	Forearm blood flow (%)	5	130	71	MD	Fixed	2.62 (2.32, 2.91)	
Kanbay(408)*	2014	General	Allopurinol	Placebo/no treatment	Endothelial-dependent dilatation (%)	11	415	213	MD	Fixed	2.69 (2.49, 2.89)	
Kanbay(408)*	2014	General	Allopurinol	Placebo/no treatment	Endothelial independent dilatation (%)	5	216	113	MD	Fixed	0.20 (-0.20, 0.61)	
Mortality												
Chaudhari(409)	2012	Infants with hypoxic-ischaemic encephalopathy	Allopurinol	Placebo/no treatment	Death during neonatal or infancy	3	114	58	RR	Fixed	0.88 (0.56, 1.38)	
Chaudhari(409)	2012	Infants with hypoxic-ischaemic encephalopathy	Allopurinol	Placebo/no treatment	Death during neonatal or infancy	2	41	20	RR	Fixed	0.97 (0.62, 1.51)	

Author	Year	Population	SUA-lowering therapy	Versus	Outcome	N studies	N participants	N Cases	Type of metric	Effect model	Reported summary effect (95%)	Concordance in overlapped MA
Chaudhari(409)	2012	Infants with hypoxic-ischaemic encephalopathy	Allopurinol	Placebo/no treatment	Death or serve neurodevelopmental disability in survivors	3	110	56	RR	Fixed	0.78 (0.56, 1.08)	
Chaudhari(409)	2012	Infants with hypoxic-ischaemic encephalopathy	Allopurinol	Placebo/no treatment	Death or serve neurodevelopmental disability in survivors	2	41	20	RR	Fixed	0.92 (0.66, 1.30)	
Other outcomes												
Chaudhari(409)	2012	Infants with hypoxic-ischaemic encephalopathy	Allopurinol	Placebo/no treatment	Severe quadriplegia in surviving infants	3	73	38	RR	Fixed	0.59 (0.28, 1.27)	
Chaudhari(409)	2012	Infants with hypoxic-ischaemic encephalopathy	Allopurinol	Placebo/no treatment	Seizures in neonatal period	3	114	58	RR	Fixed	0.93 (0.75, 1.16)	
Agarwal(410)	2013	Patients with elevated SUA or kidney dysfunction	Allopurinol	Placebo/no treatment	SBP (mmHg)	10	738	NA	MD	Random	-3.33 (-5.25, -1.42)	NO(406, 410) (discordance in statistical significance)
Bose(406)	2014	Patients with CKD or decreased kidney function	Allopurinol	Placebo/no treatment	SBP (mmHg)	5	309	NA	MD	Random	-2.70 (-7.30, 1.90)	
Agarwal(410)	2013	Patients with elevated SUA or kidney dysfunction	Allopurinol	Placebo/no treatment	DBP (mmHg)	10	738	NA	MD	Random	-1.29 (-2.48, -0.10)	NO(406, 410) (discordance in statistical significance)
Bose(406)	2014	Patients with CKD or decreased kidney function	Allopurinol	Placebo/no treatment	DBP (mmHg)	5	309	NA	MD	Random	-1.90 (-4.90, 1.20)	

Abbreviations: MA, meta-analysis; CKD, chronic kidney disease; SBP, systolic blood pressure; DBP, diastolic blood pressure; MD, mean difference; SMD, standard mean difference; NA, not available. * Quantitative analyses were not performed, because we suspected some of the reported data misused standard error as standard deviation.

Supplementary Table 3 - 5: General characteristics and main findings of the 107 Mendelian randomisation studies.*

Author	Year	Population	N/n Events (N studies)*	Outcomes	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate effect (95%CI)	P value	Concordance in overlapped MR
Anthropometric variables										
Korostishevsky(321)	2016	British	3,953	Appendicular lean mass (kg)	rs737267 in *SCL2A9*	NA	β	0.01 (NA, NA)	0.51	
Lyngdoh(322)	2012	Swiss	6,184	Fat mass (kg)	rs6855911 in *SCL2A9*	3.2%	β	0.05 (-0.10, 0.19)	0.52	
Burgess(342)	2015	European	7,158	BMI (kg/m^2)	Genetic risk score of 29 SUA-related SNPs	NA	β	-0.12(-0.53, 0.29)	0.57	NO(322, 323, 328, 329, 337, 342, 348) (discordance in direction)
Palmer(328)	2013	Danish	68,674 (2)*	BMI (kg/m^2)	rs7442295 in *SCL2A9*	2.2%	MD	-0.04 (-0.25, 0.16)	NA	
Hughes(337)	2013	European	7,979 (2)*	BMI (kg/m^2)	Genetic risk score of 5 SUA-related SNPs	2.3%	β	-0.05 (-0.12, 0.01)	0.11	
Lyngdoh(322)	2012	Swiss	6,184	BMI (kg/m^2)	rs6855911 in *SCL2A9*	3.2%	β	-0.01 (-0.16, 0.14)	0.94	
Oikonen(329)	2012	Finnish (male)	1,985	BMI (kg/m^2)	rs13129697 in *SCL2A9*	NA	β	0.04 (NA, NA)	0.82	
Parsa(348)	2012	American	868	BMI (kg/m^2)	rs16890979 in SCL2A9	NA	MD	0.24 (-0.33, 0.81)	0.39	
White(323)	2016	European	127,600 (64)*	BMI (kg/m^2)	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	-0.0003 (-0.0008, 0.0002)	NA	
Lyngdoh(322)	2012	Swiss	6,184	Waist circumference (cm)	rs6855911 in *SCL2A9*	3.2%	β	0.08 (-0.05, 0.21)	0.24	
Xiong(325)	2016	Chinese	1,667	BMD in femoral neck (g/cm^2)	Genetic risk score of 5 SUA-related SNPs	1.8%	β	0.19 (-0.42, 0.81)	0.53	NO(324, 325) (discordance in direction)
Dalbeth(324)	2015	Americans	2,501	BMD in femoral neck (g/cm^2)	Genetic risk score of 5 SUA-related SNPs	3.3%	β	-0.27 (-0.58, 0.03)	0.08	
Xiong(325)	2016	Chinese	1,667	BMD in L1–L4 (g/cm^2)	Genetic risk score of 5 SUA-related SNPs	1.8%	β	0.39 (-0.26, 0.98)	0.26	
Dalbeth(324)	2015	American	2,501	BMD in spine (g/cm^2)	Genetic risk score of 5 SUA-related SNPs	3.3%	β	0.08 (-0.32, 0.48)	0.68	
Dalbeth(324)	2015	American	2,501	BMD in total femur (g/cm^2)	Genetic risk score of 5 SUA-related SNPs	3.3%	β	-0.29 (-0.60, 0.01)	0.06	
Xiong(325)	2016	Chinese	1,667	BMD in total hip (g/cm^2)	Genetic risk score of 5 SUA-related SNPs	1.8%	β	0.19 (-0.36, 0.74)	0.50	

Author	Year	Population	N/n Events (N studies)*	Outcomes	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate effect (95%CI)	P value	Concordance in overlapped MR
Cardiovascular outcomes										
Kleber(326)	2015	German	3,060/444	Arrhythmia	Genetic risk score of 8 SUA-related SNPs	NA	OR	0.98 (0.88, 1.08)	0.64	
Kleber(326)	2015	German	3,060/368	Atrial fibrillation	Genetic risk score of 8 SUA-related SNPs	NA	OR	1.03 (0.93, 1.15)	0.57	
Kleber(326)	2015	German	3,060/316	Cardiomyopathy	Genetic risk score of 8 SUA-related SNPs	NA	OR	1.00 (0.89, 1.12)	0.93	
White(323)	2016	European	206,822/65,877 (58)*	CHD	Genetic risk score of 31 SUA-related SNPs	4.2%	OR	1.05 (0.92, 1.20)	0.49	NO(323, 327, 333, 345, 353) (discordance in direction)
Keenan(327)	2016	Pakistani	122,776/54,501 (2)*	Coronary heart disease	Genetic risk score of 14 SUA-related SNPs	3.1%	OR	1.02 (0.92, 1.12)	0.73	
Kleber(326)	2015	German	3,060/2,418	Coronary artery disease (CAD)	Genetic risk score of 8 SUA-related SNPs	NA	OR	0.99 (0.91, 1.09)	0.90	
Han(345)	2015	Chinese	2,292/1,123	CHD	rs11722228 in *SLC2A9*	NA	OR	1.09 (0.88, 1.35)	0.43	
					rs4148152 in *ABCG2*	NA	OR	0.84 (0.70, 1.11)	0.31	
Yang(333)	2010	American	23,362/3,050 (5)*	CHD incidence	Genetic risk score of 8 SUA-related SNPs	6.0%	OR	1.03 (0.85, 1.25)	0.76	
Stark(353)	2009	German	2,714/1,473	CAD	10 SUA-related SNPs:	NA	OR	Overall P>0.05		
					rs12129861			1.04 (0.93, 1.15)	0.54	
					rs780094			0.95 (0.85, 1.06)	0.39	
					rs734553			1.14 (1.00, 1.29)	0.06	
					rs734553			1.14 (0.96, 1.36)	0.13	
					rs742132			0.98 (0.87, 1.10)	0.74	
					rs1183201			0.96 (0.86, 1.07)	0.48	
					rs12356193			0.94 (0.81, 1.09)	0.42	
					rs17300741			1.08 (0.97, 1.21)	0.15	
					rs505802			1.04 (0.93, 1.17)	0.5	

Author	Year	Population	N/n Events (N studies)*	Outcomes	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate effect (95%CI)	P value	Concordance in overlapped MR
Keenan(327)	2016	Pakistani	22,926/4,526 (2)*	Heart failure	Genetic risk score of 14 SUA-related SNPs	3.1%	OR	1.07 (0.88, 1.30)	0.51	
Palmer(328)	2013	Danish	68,674/3,742 (2)*	Ischaemic heart disease	rs7442295 in SCL2A9	2.2%	HR	0.93 (0.79, 1.09)	0.38	
Kleber(326)	2015	German	3,060/2,225	Hypertension	Genetic risk score of 8 SUA-related SNPs	NA	OR	0.98 (0.90, 1.06)	0.56	
Keenan(327)	2016	Pakistani	82,091/14,779 (2)*	Ischaemic stroke	Genetic risk score of 14 SUA-related SNPs	3.1%	OR	0.99 (0.88, 1.12)	0.93	
Kleber(326)	2015	German	3,060/295	Peripheral vascular disease	Genetic risk score of 8 SUA-related SNPs	NA	OR	0.92 (0.82, 1.04)	0.18	
Kleber(326)	2015	German	3,060/538	Valve disease	Genetic risk score of 8 SUA-related SNPs	NA	OR	1.08 (0.99, 1.19)	0.10	
Yan(330)	2016	Chinese female T2DM patients	3,207	Diabetic macrovascular	Genetic risk score of 3 SUA-related SNPs	NA	OR	1.18 (1.06, 1.33)	0.004	
Oikonen(329)	2012	Finnish (male)	1,985	Carotid artery intima-media thickness (cIMT) (mm)	rs13129697 in *SCL2A9*	NA	β	<0.0001	0.99	
Mallamaci(331)	2015	Italian	449	cIMT (mm)	rs734553 in SLC2A9	NA	β	0.40 (NA, NA)	<0.001	
Mallamaci(331)	2015	Italian	449	Arterial stiffness (internal diameter) (mm)	rs734553 in SLC2A9	NA	β	0.48 (NA, NA)	0.003	

Author	Year	Population	N/n Events (N studies)*	Outcomes	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate effect (95%CI)	P value	Concordance in overlapped MR
White(323)	2016	European	89,667 (37)*	DBP (mm Hg)	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	0.005 (0.003, 0.007)	NA	NO(323, 328, 333, 337, 346, 348, 351) (discordance in direction and statistical significance)
Sedaghat(351)	2014	Dutch	5,974	DBP (mm Hg)	Genetic risk score of 30 SUA-related SNPs	4.2%	β	-0.42 (-0.72, -0.13)	0.01	
Mallamaci(346)	2014	Italian	449	DBP (mm Hg)	rs734553 in *SCL2A9*	NA	MD	NA	0.02	
Palmer(328)	2013	Danish	68,674 (2)*	DBP (mm Hg)	rs7442295 in *SCL2A9*	2.2%	MD	0.63 (-0.04, 1.29)	NA	
Hughes(337)	2013	European	7,979 (2)*	DBP (mm Hg)	Genetic risk score of 5 SUA-related SNPs	2.3%	β	-0.002 (-0.13, 0.13)	0.97	
Yang(333)	2010	American	20,699 (5)*	DBP (mm Hg)	Genetic risk score of 8 SUA-related SNPs	6.0%	β	-0.34 (-1.04, 0.35)	0.33	
Parsa(348)	2012	American	868	DBP-Clinic visit 1 (mm Hg)	rs16890979 in *SCL2A9*	NA	MD	0.52 (-0.62, 1.66)	0.36	
Parsa(348)	2012	American	868	DBP-High-salt 24-h (mm Hg)	rs16890979 in *SCL2A9*	NA	MD	0.42 (-0.56, 1.40)	0.41	
Parsa(348)	2012	American	868	DBP-Low-salt 24-h (mm Hg)	rs16890979 in *SCL2A9*	NA	MD	0.19 (-0.75, 1.13)	0.69	
Parsa(348)	2012	American	868	DBP-Salt sensitivity 24-h (mm Hg)	rs16890979 in *SCL2A9*	NA	MD	-0.01 (-0.62, 0.60)	0.99	
White(323)	2016	European	89,667 (37)*	SBP (mm Hg)	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	0.005 (0.003, 0.006)	NA	NO(323, 328, 333, 337, 346, 348, 351) (discordance in direction and statistical significance)
Sedaghat(351)	2014	Dutch	5,974	SBP (mm Hg)	Genetic risk score of 30 SUA-related SNPs	4.2%	β	-0.75 (-1.31, -0.19)	0.01	
Mallamaci(346)	2014	Italian	449	SBP (mm Hg)	rs734553 in SLC2A9	NA	β	NA	0.02	
Palmer(328)	2013	Danish	68,674 (2)*	SBP (mm Hg)	rs7442295 in *SCL2A9*	2.2%	MD	0.65 (-0.54, 1.85)	NA	
Hughes(337)	2013	European	7,979 (2)*	SBP (mm Hg)	Genetic risk score of 5 SUA-related SNPs	2.3%	β	0.07 (-0.12, 0.26)	0.47	
Yang(333)	2010	American	20,673 (5)*	SBP (mm Hg)	Genetic risk score of 8 SUA-related SNPs	6.0%	β	-0.83 (-1.96, 0.30)	0.15	
Parsa(348)	2012	American	868	SBP-Clinic visit 1 (mm Hg)	rs16890979 in *SCL2A9*	NA	MD	0.08 (-1.70, 1.86)	0.38	
Parsa(348)	2012	American	868	SBP-High-salt 24-h (mm Hg)	rs16890979 in *SCL2A9*	NA	MD	2.20 (0.65, 3.75)	0.01	

Author	Year	Population	N/n Events (N studies)*	Outcomes	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate effect (95%CI)	P value	Concordance in overlapped MR
Parsa(348)	2012	American	868	SBP-Low-salt 24-h (mm Hg)	rs16890979 in *SCL2A9*	NA	MD	1.48 (0.09, 2.87)	0.04	
Parsa(348)	2012	American	868	SBP-Salt sensitivity 24-h (mm Hg)	rs16890979 in *SCL2A9*	NA	MD	0.62 (-0.34, 1.58)	0.21	
Metabolic disorders										
White(323)	2016	European	84,638/15,360 (20)*	Diabetes	Genetic risk score of 31 SUA-related SNPs	4.2%	OR	0.99 (0.99, 1.01)	0.82	NO(323, 326, 332) (327, 349, 354) (discordance in direction and statistical significance)
Kleber(326)	2015	German	3,060/1,236	Diabetes	Genetic risk score of 8 SUA-related SNPs	NA	OR	0.94 (0.88, 1.01)	0.10	
Sluijs(332)	2015	European	165,482/41,508 (2)*	Diabetes	Genetic risk score of 24 SUA-related SNPs	4.0%	OR	0.99 (0.92, 1.06)	0.79	
Keenan(327)	2016	Pakistani	110,452/26,488 (2)*	T2DM	Genetic risk score of 14 SUA-related SNPs	3.1%	OR	0.95 (0.86, 1.05)	0.28	
Pfister(349)	2012	British	16,064/7,504 (4)*	T2DM	Genetic risk score of 8 SUA-related SNPs	NA	OR	0.99 (0.94, 1.04)	0.68	
Sun(354)	2015	Chinese	5,198/2,999	T2DM	15 SUA-related SNPs	NA	OR	No overall result		
					rs12129861			1.03 (0.92, 1.16)	0.59	
					rs780094			1.22 (1.11, 1.35)	3.9E-05	
					rs2544390			0.97 (0.88, 1.06)	0.50	
					rs11722228			1.01 (0.91, 1.12)	0.83	
					rs16890979			0.98 (0.66, 1.45)	0.91	
					rs3775948			1.03 (0.93, 1.13)	0.59	
					rs10489070			0.96 (0.84, 1.10)	0.54	
					rs2231142			0.94 (0.85, 1.04)	0.21	
					rs742132			1.01 (0.91, 1.13)	0.80	
					rs1183201			0.98 (0.86, 1.11)	0.75	
					rs1165205			0.98 (0.87, 1.11)	0.75	
					rs1333049			1.024 (0.93, 1.12)	0.61	
					rs17300741			0.97 (0.80, 1.19)	0.78	
					rs506338			0.99 (0.89, 1.10)	0.87	
					rs606458			1.11 (1.01, 1.22)	0.04	

Author	Year	Population	N/n Events (N studies)*	Outcomes	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate effect (95%CI)	P value	Concordance in overlapped MR
Yang(333)	2010	American	25,877 (5)*	Fasting glucose (mmol/L)	Genetic risk score of 8 SUA-related SNPs	6.0%	β	-0.06 (-0.13, 0.02)	0.13	NO(323, 333, 348) (discordance in direction)
White(323)	2016	European	57,397 (28)*	Fasting glucose (mmol/L)	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	-0.001 (-0.003, 0.001)	NA	
Parsa(348)	2012	American	868	Glucose (mmol/L)	rs16890979 in *SCL2A9*	NA	β	0.78 (-0.87, 2.43)	0.36	
Yang(333)	2010	American	19,899 (5)*	Fasting insulin†	Genetic risk score of 8 SUA-related SNPs	6.0%	Z statistics	-0.015 (NA, NA)	0.99	
Dai(334)	2013	Chinese	7,827	Metabolic syndrome	Genetic risk score of 2 SNPs (*SLC2A9* and *ABCG2*)	2.1%	OR	1.03 (0.98, 1.09)	0.23	NO(334, 347) (discordance in statistical significance)
McKeigue(347)	2010	Scottish	1,017/203	Metabolic syndrome	Genetic risk score of 6 SNPs in *SCL2A9*	NA	NA	NA	>0.05	
Kidney disorders										
Yang(333)	2010	American	23,387/3,092 (5)*	CKD	Genetic risk score of 8 SUA-related SNPs	6.0%	OR	1.20 (0.96, 1.50)	0.12	
Greenberg(335)	2015	American	7,553/823	Acute kidney injury	Genetic risk score of 8 SUA-related SNPs	6.0%	HR	1.01 (0.77, 1.34)	0.92	
Testa(336)	2014	Italian	755/244	Renal events	rs734553 in *GLUT9*	NA	HR	2.35 (1.25, 4.42)	0.01	
Hughes(337)	2013	European	7,979 (2)*	eGFR (mL/min/1.73 m^2)	Genetic risk score of 5 SUA-related SNPs	2.3%	β	12.20 (-11.50, 35.90)	0.31	NO(333, 337, 348, 355, 356) (discordance in direction and statistical significance)
Parsa(348)	2012	American	868	eGFR (mL/min/1.73 m^2)	rs16890979 in *SCL2A9*	NA	MD	0.42 (-1.78, 2.62)	0.71	
Yang(333)	2010	American	23,844 (5)*	Log eGFR (mL/min/1.73 m^2)	Genetic risk score of 8 SUA-related SNPs	6.0%	β	0.001 (-0.01, 0.02)	0.91	
Tabara(355)	2010	Japanese	5,165 (2)*	eGFR (mL/min/1.73 m^2)	Genetic risk score of 3 SUA-related SNPs	NA	β	-0.11 (NA, NA)	<0.001	
Voruganti(356)	2014	Native American	3,604 (3)*	eGFR (mL/min/1.73 m^2)	7 SNPs in *SCL2A9:*	5.3%	Residual variance¶	No overall result		
					rs16890979			0.52 (NA, NA)	0.002	
					rs6832439			0.52 (NA, NA)	0.002	
					rs6449213			0.22 (NA, NA)	0.08	
					rs13131257			0.58 (NA, NA)	0.001	
					rs737267			0.44 (NA, NA)	0.004	
					rs10805346			0.69 (NA, NA)	<0.001	
					rs12498956			0.24 (NA, NA)	0.05	

Author	Year	Population	N/n Events (N studies)*	Outcomes	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate effect (95%CI)	P value	Concordance in overlapped MR
Voruganti(356)	2014	Native American	3,604 (3)*	SCr (mmol/L)	7 SNPs in *SCL2A9:*	5.3%	Residual variance¶	No overall result		NO(337, 356) (discordance in statistical significance)
					rs16890979			0.53 (NA, NA)	0.001	
					rs6832439			0.54 (NA, NA)	0.002	
					rs6449213			0.27 (NA, NA)	0.003	
					rs13131257			0.60 (NA, NA)	0.000	
					rs737267			0.46 (NA, NA)	0.004	
					rs10805346			0.68 (NA, NA)	0.000	
					rs12498956			0.23 (NA, NA)	0.06	
Hughes(337)	2013	European	7,979 (2)*	SCr (mmol/L)	Genetic risk score of 5 SUA-related SNPs	2.3%	β	-19.23 (-40.32, 1.86)	0.07	
Voruganti(356)	2014	Native American	3,604 (3)*	Albumin/creatinine ratio	7 SNPs in *SCL2A9:*	5.3%	Residual variance¶	Overall P>0.05		
					rs16890979			0.13 (NA, NA)	0.07	
					rs6832439			0.16 (NA, NA)	0.05	
					rs6449213			0.01 (NA, NA)	0.64	
					rs13131257			0.14 (NA, NA)	0.05	
					rs737267			0.14 (NA, NA)	0.06	
					rs10805346			0.14 (NA, NA)	0.10	
					rs12498956			0.07 (NA, NA)	0.15	
All-cause and cause-specific mortality										
Kleber(326)	2015	German	3,060/na	Cardiovascular mortality	Genetic risk score of 8 SUA-related SNPs	NA	aHR	1.11 (1.02, 1.21)	0.02	
Kleber(326)	2015	German	3,060/na	All-cause mortality	Genetic risk score of 8 SUA-related SNPs	NA	aHR	1.02 (0.95, 1.09)	0.59	

Author	Year	Population	N/n Events (N studies)*	Outcomes	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate effect (95%CI)	P value	Concordance in overlapped MR
Kleber(326)	2015	German	3,060/na	Sudden cardiac death	Genetic risk score of 8 SUA-related SNPs	NA	aHR	1.18 (1.03, 1.35)	0.02	
Metabolites										
White(323)	2016	European	196,621 (68)*	HDL-C (mmol/L)	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	-0.008 (-0.010, -0.006)	NA	
White(323)	2016	European	196,621 (68)*	LDL-C (mmol/L)	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	-0.001 (-0.003, 0.001)	NA	
White(323)	2016	European	196,621 (68)*	TC (mmol/L)	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	0.000 (-0.002, 0.002)	NA	
White(323)	2016	European	196,621 (68)*	TG (mmol/L)	Genetic risk score of 31 SUA-related SNPs	4.2%	MD§	0.014 (0.013, 0.016)	NA	NO(323, 348, 350) (discordance in direction)
Rasheed(350)	2014	European	8,208 (2)*	TG (mmol/L)	Genetic risk score of 5 SUA-related SNPs	1.7%	β	−1.01 (-2.57, 0.56)	0.21	
Parsa(348)	2012	American	868	TG (mmol/L)	rs16890979 in *SCL2A9*	NA	MD	2.38 (-2.87, 7.63)	0.35	
Xiong(325)	2016	Chinese	1,667	Parathyroid hormone (pg/mL)	Genetic risk score of 5 SUA-related SNPs	1.8%	β	-0.63 (-2.12, 0.85)	0.40	
Xiong(325)	2016	Chinese	1,667	Phosphorus (mmol/L)	Genetic risk score of 5 SUA-related SNPs	1.8%	β	-0.16 (-0.74, 0.42)	0.59	
Burgess(342)	2015	European	7,158	C-reactive protein (CRP) (mg/L)	Genetic risk score of 29 SUA-related SNPs	NA	β	-0.05 (-0.15, 0.05)	0.37	
Xiong(325)	2016	Chinese	1,667	Calcium (mmol/L)	Genetic risk score of 5 SUA-related SNPs	1.8%	β	0.06 (-0.10, 0.21)	0.48	
Xiong(325)	2016	Chinese	1,667	Tropocollagen type 1 N-terminal propeptide (ng/L)	Genetic risk score of 5 SUA-related SNPs	1.8%	β	0.11 (-1.53, 1.75)	0.90	
Xiong(325)	2016	Chinese	1,667	β-crosslaps of type I collagen (ng/L)	Genetic risk score of 5 SUA-related SNPs	1.8%	β	-1.45 (-3.17, 0.27)	0.10	
Xiong(325)	2016	Chinese	1,667	25(OH)D (ng/mL)	Genetic risk score of 5 SUA-related SNPs	1.8%	β	0.76 (-0.63, 2.15)	0.28	

Author	Year	Population	N/n Events (N studies)*	Outcomes	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate effect (95%CI)	P value	Concordance in overlapped MR
Neurocognitive disorders										
Gonzalez-Aramburu(344)	2014	Spanish	343	Dementia in PD	Genetic risk score of 8 SUA-related SNPs	NA	OR	1.05 (0.70, 3.00)	0.31	NO(338, 339, 343, 344, 352) (discordance in direction and statistical significance)
Simon(352)	2014	American	808 (2)*	PD progression	Genetic risk score of 3 SNPs in *SCL2A9*	NA	HR	1.16 (1.00, 1.35)	0.06	
Gao(338)	2013	American	1,699	Parkinson's disease	12 SNPs in *SCL2A9*	NA	OR	Overall P>0.05		
					rs16890979			1.06 (0.90, 1.24)	0.51	
					rs13129697			0.99 (0.85, 1.16)	0.91	
					rs737267			1.01 (0.86, 1.18)	0.91	
					rs6855911			1.00 (0.85, 1.18)	0.98	
					rs4697700			1.03 (0.88, 1.22)	0.69	
					rs4481233			1.01 (0.85, 1.21)	0.90	
					rs7442295			1.06 (0.90, 1.26)	0.50	
					rs6449213			1.01 (0.85, 1.21)	0.88	
					rs1014290			1.04 (0.88, 1.22)	0.67	
					rs12509955			0.99 (0.84, 1.18)	0.93	
					rs17251963			1.05 (0.88, 1.25)	0.59	
					rs12510549			0.96 (0.81, 1.14)	0.65	
Gonzalez-Aramburu(343)	2013	British	1,061	Parkinson's disease	Genetic risk score of 9 SUA-related SNPs	NA	OR	1.55 (1.10, 2.18)	0.01	
Facheris(339)	2011	European	664 (3)*	Age at onset of PD	4 SNPs in *SCL2A9*	NA	β	Null after multiple testing correction		
					rs737267	NA		3.10 (0.17, 6.03)	0.04	
					rs6449213	NA		-1.18 (-4.96, 2.59)	0.54	
					rs1014290	NA		-4.56 (-8.13, -1.00)	0.01	
					rs733175	NA		3.59 (0.67, 6.51)	0.02	
Lyngdoh (340)	2013	Swiss	3,716/660	Lifetime anxiety disorders	rs6855911 in SLC2A9	3.2%	OR (male)	1.40 (1.07, 1.84)	0.02	
							OR (female)	0.97 (0.80, 1.17)	0.73	

Author	Year	Population	N/n Events (N studies)*	Outcomes	Genetic instruments (GI)	SUA variance (R^2) explained by GI	Type of metric	Estimate effect (95%CI)	P value	Concordance in overlapped MR
Lyngdoh(340)	2013	Swiss	3,716/370	Current anxiety disorders	rs6855911 in SLC2A9	3.2%	OR (male)	1.42 (0.99, 2.03)	0.06	
							OR (female)	0.84 (0.66, 1.06)	0.14	
Houlihan(341)	2010	Scottish (Population 1: LBC1936)	1,091	Memory performance	4 SNPs in *SCL2A9*	NA	β	Overall P<0.05		NO(341) (discordance in statistical significance)
					rs733175			-0.10 (NA, NA)	0.0002	
					rs1014290			-0.07 (NA, NA)	0.01	
					rs6449213			-0.07 (NA, NA)	0.01	
					rs6449213			-0.07 (NA, NA)	0.01	
Houlihan(341)	2010	Scottish (Population 2: ET2DS)	1,066	Memory performance	4 SNPs in *SCL2A9*	NA	β	Overall P>0.05		
					rs733175			-0.03 (NA, NA)	0.27	
					rs1014290			-0.04 (NA, NA)	0.22	
					rs6449213			-0.03 (NA, NA)	0.41	
					rs6449213			-0.04 (NA, NA)	0.19	
Other outcomes										
Kleber(326)	2015	German	3,060/226	Cancer	Genetic risk score of 8 SUA-related SNPs	NA	OR	0.95 (0.83, 1.08)	0.41	
Keenan(327)	2016	Pakistani	71,501/3,151 (2)*	Gout	Genetic risk score of 14 SUA-related SNPs	3.1%	OR	5.84 (4.56,0 7.49)	3.55E-40	NO(326, 327, 333) (discordance in statistical significance)
Kleber(326)	2015	German	3,060/19	Gout	Genetic risk score of 8 SUA-related SNPs	NA	OR	1.15 (0.72, 1.82)	0.56	
Yang(333)	2010	American	25,982/1,033 (5)*	Gout	Genetic risk score of 8 SUA-related SNPs	6.0%	OR	12.40 (8.50, 18.00)	3.00E-39	

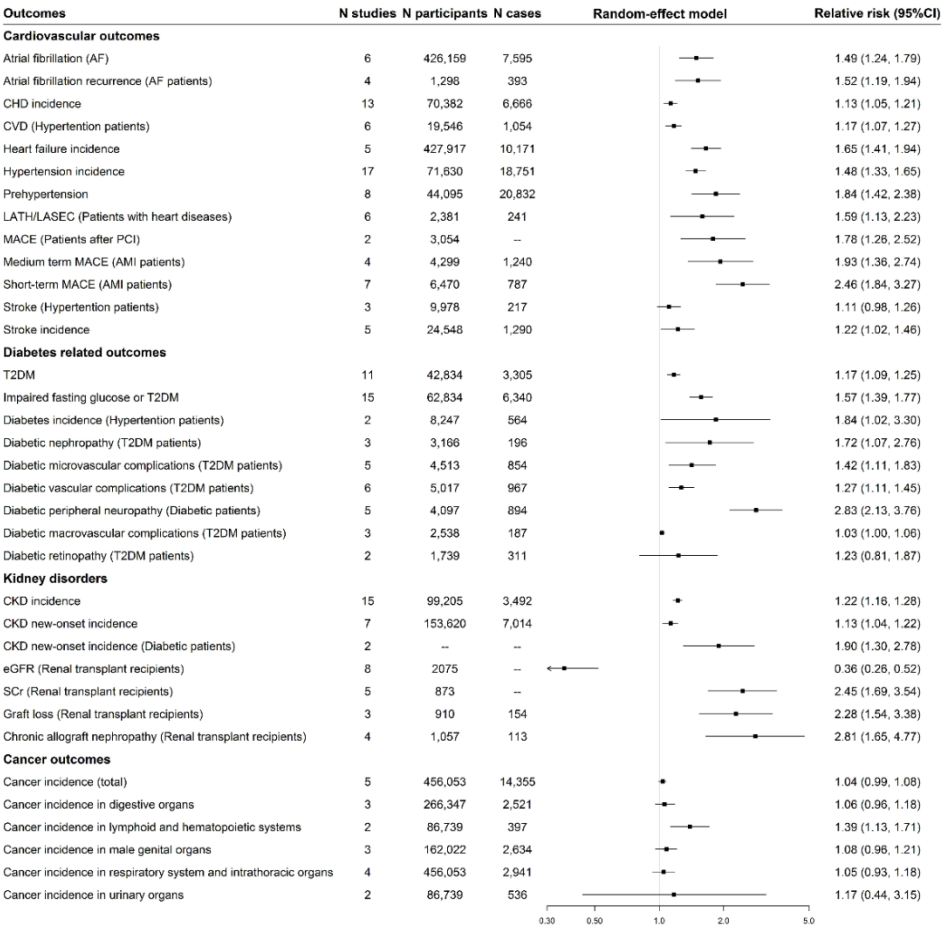
Abbreviations: MR, Mendelian randomisation study; BMI, body mass index; BMD, bone mineral density; SBP, systolic blood pressure; DBP, diastolic blood pressure; CHD, coronary heart disease; IMT, intima-media thickness; CVD, cardiovascular disease; T2DM, Type 2 diabetes; CKD, chronic kidney disease; SCr, serum creatinine; eGFR, glomerular filtration rate; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TG, triglyceride; TC, total cholesterol; PD, Parkinson's disease; LBC1936, the Lothian birth cohort; ET2DS, the Edinburgh type 2 diabetes study; MD, mean difference; NA, not available.

* If the outcomes were reported from Mendelian randomisation analysis with pooling multiple studies, the number of studies included in pooled analysis was displayed in brackets.

† Because of the lack of a standard to covert insulin in different studies to the same scale, sample size-weighted pooled meta-analyses were performed and Z statistics were reported instead of the β coefficient.

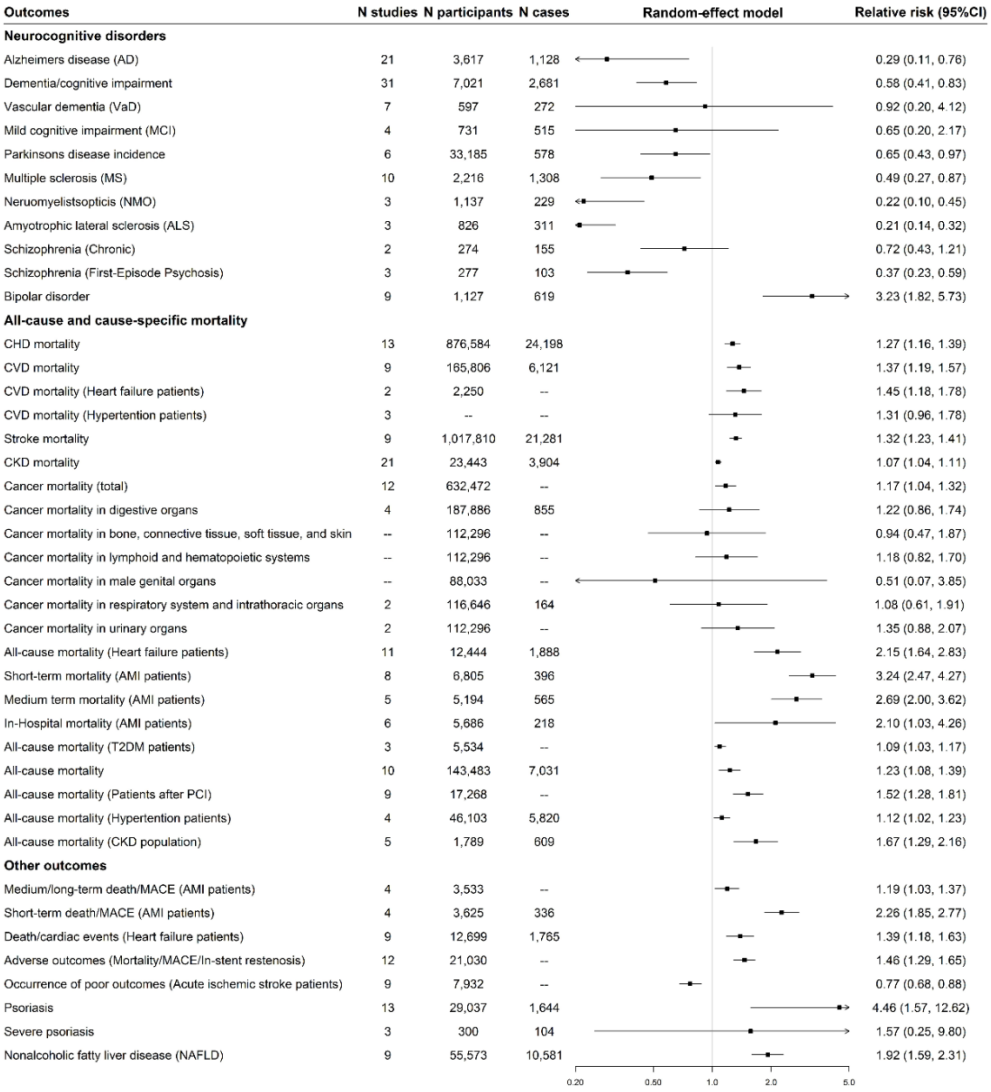
§ MD (mean difference) represented the difference in mean caused by per inverse variance weighted allele estimated from meta-analyses.

¶ Residual variance represented the proportion of residual variance explained by the SUA related SNPs.



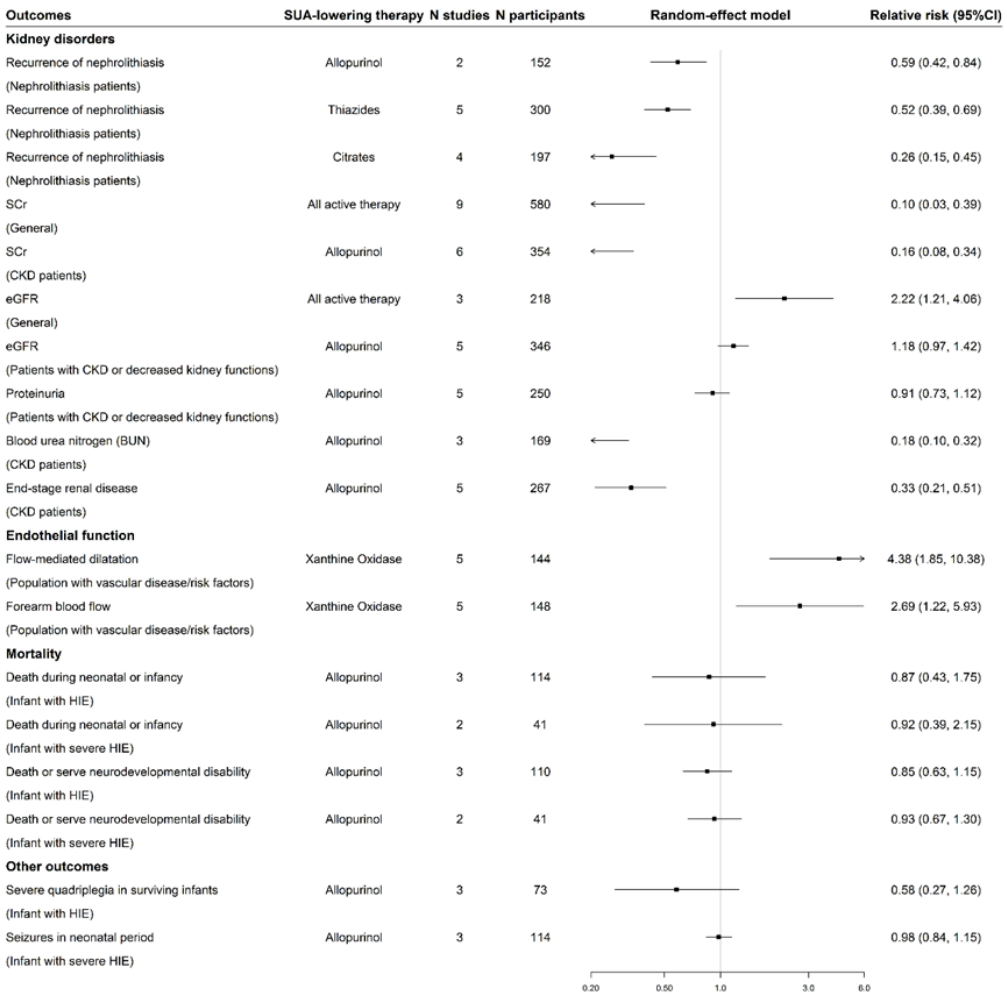
Supplementary Figure 3 - 1: Summary random-effect estimates of cardiovascular, diabetes, kidney disorders and cancer outcomes reported in meta-analyses of observational studies.

Abbreviations: AF, atrial fibrillation; CHD, coronary heart disease; CVD, cardiovascular disease; LATH/LASEC, left atrial thrombus or spontaneous echo contrast; PCI, percutaneous coronary intervention; MACE, major adverse cardiovascular events; AMI, acute myocardial infarction; T2DM, type 2 diabetes; CKD, chronic kidney disease; SCr, serum creatinine; eGFR, glomerular filtration rate.



Supplementary Figure 3 - 2: Summary random-effect estimates of neurocognitive disorders, all-cause and cause-specific mortality, and other outcomes reported in meta-analyses of observational studies.

Abbreviations: CHD, coronary heart disease; CVD, cardiovascular disease; T2DM, type 2 diabetes; MACE, major adverse cardiovascular events; AMI, acute myocardial infarction; PCI, percutaneous coronary intervention; CKD, chronic kidney disease; NAFLD, non-alcoholic fatty liver disease.



Supplementary Figure 3 - 3: Summary random-effect estimates of health outcomes reported in meta-analyses of RCTs.

Abbreviations: CKD, chronic kidney disease; SCr, serum creatinine; eGFR, glomerular filtration rate; HIE, hypoxic-ischaemic encephalopathy.

4 DESCRIPTION AND MANIPULATION OF UK BIOBANK DATABASE

This chapter describes the UK Biobank resource used in this thesis. The first section of this chapter presents the study design, ethical approval, participant recruitment, data collection and data release of UK Biobank cohort. The second part provides a summary description of the UK Biobank dataset. The third part presents the processes of data cleaning and data preparation and summarises the characteristics of the datasets used in the following analysis.

4.1 The UK Biobank cohort

4.1.1 Study design

The UK biobank is a large-scale, population-based prospective cohort study designed to improve the prevention, diagnosis and treatment of a wide range of diseases. It was funded by a body of organisations including the Wellcome Trust, Medical Research Council, Department of Health, Scottish Government, Northwest Regional Development Agency, Welsh Government, British Heart Foundation and Diabetes UK. This cohort was designed to include ~500,000 participants and to combine extensive measurements of baseline data and genetic data with longitudinal follow-up of participants' national medical records (e.g. in-patient hospital episode records and data from the cancer registry and death registry). The participant recruitment and baseline data collection took place from 2006 to 2010. The longitudinal follow-up will last for 20 years to allow detailed investigation of the genetic and non-genetic determinants of a wide range of complex diseases and phenotypes. The establishment of the cohort, the baseline assessment, the measurement and quality control of the genotype data, and the collection of the medical records for longitudinal follow-up were all carried out centrally by the UK Biobank team. Further manipulation of the UK Biobank data, for example, selecting study population, constructing genetic instruments and defining the phenome framework were performed by myself with help from collaborators.

4.1.1.1 Ethics approval and research ethics requirements

The detailed research protocol and ethics and governance aspects of the UK Biobank project have gone through an extensive review. An independent Ethics and Governance Council (EGC) was established by the MRC and the Wellcome Trust to ensure that the UK Biobank project met the required standards for conducting research on human participants. The key ethics and governance principles of UK Biobank were presented in the Ethics and

Governance Framework (EGF) (411). Based on the standards set by the EGF, research activities of UK Biobank were approved by the North West Multi-Centre Research Ethics Committee (MREC) in relation to the process of participant invitation and assessment and follow-up procedures. Additionally, ethics approvals from the National Information Governance Board for Health & Social Care (NIGB) in England and Wales and approval from the Community Health Index Advisory Group (CHIAG) in Scotland were also obtained to gain access to the information that would allow invitation of participants. UK Biobank had also sought a generic Research Tissue Bank (RTB) approval, which covered the vast majority of research using this resource, instead of requiring each application to apply for separate ethics approval. Informed consent was given by participants during their visit to the assessment centres. This consent related to their understanding and awareness of the following aspects: the purpose of UK Biobank, the information and samples that will be collected at enrolment, the linkage to their full medical records, the role of UK Biobank as the legal owner of the datasets, the safeguards in place relating to data and samples, the possibility of being re-contacted and the right to withdraw at any time without giving any reason and without penalty.

The research protocol of this study was reviewed by the UK Biobank committee to ensure this study was consistent with the access procedures, the EGF and the consent provided by the participants. The application (application ID: 10775) was officially approved by the UK Biobank committee in 2015. The study did not require to re-contact the participants and did not involve any use of samples that were not covered by the RTB approval; therefore, following the instruction from the National Research Ethics Service (NRES) and UK Biobank's governing Research Ethics Committee (REC), a separate ethics approval was not required for this study. When performing the data analysis, I complied with the UK Biobank Access Policy and the EGF regulation (411, 412) and acted in accordance with the Data Protection Act (DPA) (413). Findings deriving from the UK Biobank resource have been and will be published with the approval from UK Biobank. Knowledge developed from this study will be disseminated to benefit public health.

4.1.1.2 Participant recruitment and enrolment

Potentially eligible participants of UK Biobank were identified from the National Health Service (NHS) patient registry (414). People who were registered with the UK NHS, between the ages of 40 and 69 years old and living within 25 miles from any of the local study assessment centres, were eligible to participate in the study. The assessment centres were located across the UK and included Edinburgh, Glasgow, Newcastle, Middlesbrough,

Leeds, Sheffield, Bury, Manchester, Liverpool, Wrexham, Stoke, Nottingham, Birmingham, Oxford, Reading, Bristol, Swansea, Hounslow, Central London, and Croydon (**Figure 4-1**). The NHS number and date of birth from NHS register data were used to verify the age (40-69 years old) of potential participants and remove duplicates or death records. A list of contact details of eligible participants was generated from the NHS register data by stratifying key demographic characteristics (e.g., age, gender and postcode as an index of social deprivation) to recruit a widely generalisable population. Over-sampling was performed for several particular diseases of interest (e.g. chronic obstructive pulmonary disease [COPD], schizophrenia) in order to recruit an adequate number of cases.

Figure 4 - 1: The geographic locations of assessment centres across UK.
(Source: adapted from (414)).

Potential participants were sent an invitation letter for participation. They were asked to confirm the pre-booked provisional appointment at the local assessment centre. People who didn't want to participate in the study were encouraged to indicate their unwillingness and to cancel the appointment. People who confirmed the appointment were sent a written confirmation of their appointment details along with instructions on preparing for the baseline assessment. A pre-visit reminder message or mail was sent to the confirmed participants before the scheduled appointment for baseline assessment. Between 2006 and

2010, about 9,000,000 people were invited by mail and 502,656 individuals were finally enrolled. The median number of participants recruited across the 23 assessment centres was 21,290 (range: 649 to 44,220) (**Table 4-1**).

Table 4 - 1: The number of participants recruited across the 23 assessment centres.

Clinic ID	Assessment centre	Dates of operation	No. of recruitment
11021	Birmingham	29/10/2009 - 21/07/2010	25,503
11011	Bristol	09/07/2008 - 28/11/2009	43,015
11008	Bury	14/01/2008 - 20/12/2008	28,336
11003	Cardiff	08/10/2007 - 31/05/2008	17,882
11024	Cheadle (revisit)	01/08/2012 - 06/06/2013	20,346
11020	Croydon	24/09/2009 - 09/07/2010	27,385
11005	Edinburgh	07/11/2007 - 07/06/2008	17,201
11004	Glasgow	16/07/2007 - 19/04/2008	18,651
11018	Hounslow	17/06/2009 - 26/06/2010	28,879
11010	Leeds	27/02/2008 - 11/07/2009	44,209
11016	Liverpool	28/01/2009 - 01/04/2010	32,818
11012	London Barts	27/08/2008 - 29/08/2009	12,583
11001	Manchester	16/04/2007 - 22/12/2007	13,940
11017	Middlesbrough	29/04/2009 - 06/02/2010	21,289
11009	Newcastle	23/01/2008 - 28/03/2009	37,008
11013	Nottingham	30/07/2008 - 12/09/2009	33,877
11002	Oxford	30/04/2007 - 27/10/2007	14,062
11007	Reading	14/05/2008 - 02/05/2009	29,417
11014	Sheffield	05/08/2009 - 13/07/2010	30,397
10003	Stockport (pilot)	13/03/2006 - 13/06/2006	3,798
11006	Stoke	05/12/2007 - 26/07/2008	19,440
11022	Swansea	11/03/2010 - 03/07/2010	2,281
11023	Wrexham	16/08/2010 - 01/10/2010	649

(Source: adapted from (414)).

4.1.2 Baseline assessment and data collection

4.1.2.1 Overview of the baseline assessment process

When individuals attended the assessment visit, UK Biobank staff provided an explanation and clarification about the research process. They were asked for their consent to participate

and went through a series of assessment stations. People who were unable to give consent or unable to take part in data collection or who were uncomfortable with any aspect of the participation were not enrolled. In general, information on participants' sociodemographic status, family history, early life exposure, lifestyle and environmental exposures, health status, and psychosocial factors were collected by a self-completed questionnaire and computer-assisted personal interview. A series of physical and functional measurements were also taken for further analysis (e.g., anthropometrics, blood pressure and heart rate, spirometry, and eye examinations). Biological samples of blood and urine were collected for biochemical tests. The process of baseline assessment is summarised in **Figure 4-2**.

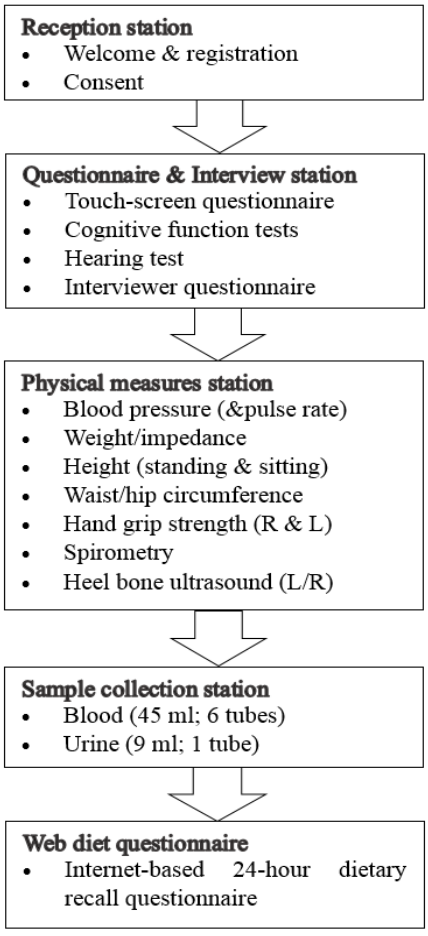


Figure 4 - 2: Baseline assessment process.

(Source: adapted from (414)).

4.1.2.2 Baseline data collected from the questionnaire and interview

The UK Biobank questionnaire collected a variety of baseline data, which could be summarised into the following broad data fields (**Table 4-2**): sociodemographic factors,

lifestyle and environmental exposures, family history and early life exposures, general health and disability, psychiatric or psychosocial state, and cognitive function (415).

- ***Sociodemographic factors:*** Sociodemographic information on education and qualifications, employment status and current occupation, marital status, car ownership, household income, household structure, ethnicity and country of birth, was collected during the baseline assessment.
- ***Environmental factors:*** A large number of environmental exposures were collected for the UK Biobank participants, including living address, residence at birth, occupation and other workplace factors, sleep, domestic heating, indoor air pollution and mobile phone use, *etc*.
- ***Smoking and alcohol:*** Comprehensive questions on smoking were asked for those who smoked; alcohol consumption was assessed in terms of quantity, frequency and beverage specificity.
- ***Physical activity:*** Physical activity was assessed by a self-ranking of activity level (vigorous, moderate and walking) and a 24-hour recall of daily activities.
- ***Dietary habits:*** Dietary habits were recorded by a self-administered questionnaire developed based on the European standardised program for computer-assisted 24-hour dietary recall instruments (EPIC-SOFT).
- ***Family history and early life exposures:*** Family history of common serious illnesses and early life exposures on birth weight, maternal smoking, breastfeeding, and childhood body size were collected during the baseline assessment.
- ***General health and disability:*** Data on medical conditions, general health questions, self-reported disability, wheeze, chronic pain and chest pain, reproductive history of women were all collected.
- ***Psychosocial and psychiatric state:*** Psychological and psychiatric traits were assessed by a series of standardised questionnaires (neuroticism/mood).
- ***Cognitive function:*** Paired-associated learning questions to assess global cognition and reaction time tests for touch-screen administration were adopted to assess the cognitive function of the study participants.

Table 4 - 2: A summary of data collected from questionnaire and interview.

Data fields	Variables available
Sociodemographic factors	Social class, education and qualifications, employment status and current occupation, marital status, car ownership, household income, household structure, ethnicity and country of birth, *etc*.
Family history and early life exposures	Family history of illness, birth weight, breast feeding, maternal smoking, childhood body size/height, place of birth, being a twin or other multiple order birth, *etc*.
Lifestyle and environmental factors	Physical activity, smoking, diet, alcohol consumption, sleep, occupations, domestic heating, *etc*.
General health and disability	Medical conditions, medications, disability, hearing, sight, reproductive history, chronic pain and chest pain, wheeze, skin and hair colour, *etc*.
Psychiatric or psychosocial factors	Neurosis, mood, depression status, satisfaction (job, family, health), social support, mental categories, and history of psychiatric care, *etc*.
Cognitive function	Fluid intelligence, numeric/prospective memory, pairs matching, reaction time, trail making, *etc.*

4.1.2.3 Physical measurements taken at baseline assessment

Baseline data from physical measurements were taken by well-trained stuff during the participants' visit to the assessment centre (**Table 4-3**). These included:

- ***Blood pressure (and pulse rate):*** Blood pressure and pulse rate were measured twice by the Omron HEM-7015IT digital blood pressure monitor.
- ***Weight:*** Weight was measured by the Tanita BC-418 MA body composition analyser with removing shoes and heavy outer clothing.
- ***Height:*** Standing and sitting heights (shoeless) were measured using a Seca 202 height measure.
- ***Waist and hip circumference***: Waist and hip circumferences were measured using a Wessex non-stretchable sprung tape measure.
- ***Bio-impedance:*** Bio-impedance was measured by the Tanita BC-418MA body composition analyser.
- ***Hand grip strength:*** Right- and left-hand grip strengths were measured once each using a Jamar J00105 hydraulic hand dynamometer.
- ***Spirometry:*** The forced expiratory volume in 1 second (FEV1) and forced vital capacity (FVC) were measured three times using the Vitalograph Pneumotrac 6800 spirometer.
- ***Bone densitometry:*** Bone mineral density (left heel) was measured by the Norland

McCue Contact Ultrasound Bone Analyser (CUBA).

- ***Other measurements:*** Measurements (e.g., electrocardiogram [ECG], ankle-brachial index, pulse wave velocity, carotid intimal-medial thickness) were available in the pilot study but excluded from the main assessment.

Table 4 - 3：A summary of data collected from physical measurements.

Data fields	List of variables
Anthropometrics	Height, weight, bio-impendence, hip and waist circumference, *etc.*
Blood pressure and heart rate	Systolic blood pressure, diastolic blood pressure, peripheral pulse pressure, average heart rate, *etc.*
Arterial stiffness	Pulse wave reflection index, pulse wave pressure versus time response curve, pulse wave arterial stiffness index, *etc.*
Spirometry	Forced vital capacity (FVC), forced expiratory volume in 1-second (FEV1), peak expiratory flow (PEF), *etc.*
Eye examination	Visual acuity, refractive index, intra-ocular pressure, optical-coherence tomography, *etc.*
Others	Grip strength, bone mineral density, electrocardiograph, *etc.*

4.1.3 Genotypic data and quality control (QC)

4.1.3.1 DNA extraction

The blood samples collected from participants at the baseline assessment were stored in the UK biobank facilities (at either -80°C or -196°C) in Stockport, UK. Full details of the procedures of sample retrieval and DNA extraction are provided in the following references (416, 417). Briefly, buffy coat samples for genotyping were located by robot to a 96-position plate; then DNA was extracted and purified by an automated process system using the Maxwell 16 Instrument. The Maxwell 16 Blood DNA Purification Kit was applied to test if the extracted DNA samples were free from contamination and had suitable concentration and purity. The DNA quantification was assessed with a set of pre-defined criteria: (i) average plate concentration was at least 20 ng/μl; (ii) average absorbance ratio (measured at 260 nm / 280nm) was between 1.8-2.2; (iii) at least 80% of sample DNA concentrations was greater than 10 ng/μl; (iv) at least 80% of sample absorbance ratio (measured at 260 nm / 280nm) was between 1.8-2.2; (v) the extraction concentration of negative control was less than 1 ng/μl. Samples failing to meet the pre-defined criteria (if possible these samples were re-processed) were excluded. A brief overview of the process for DNA extraction is shown in **Figure 4-3**.

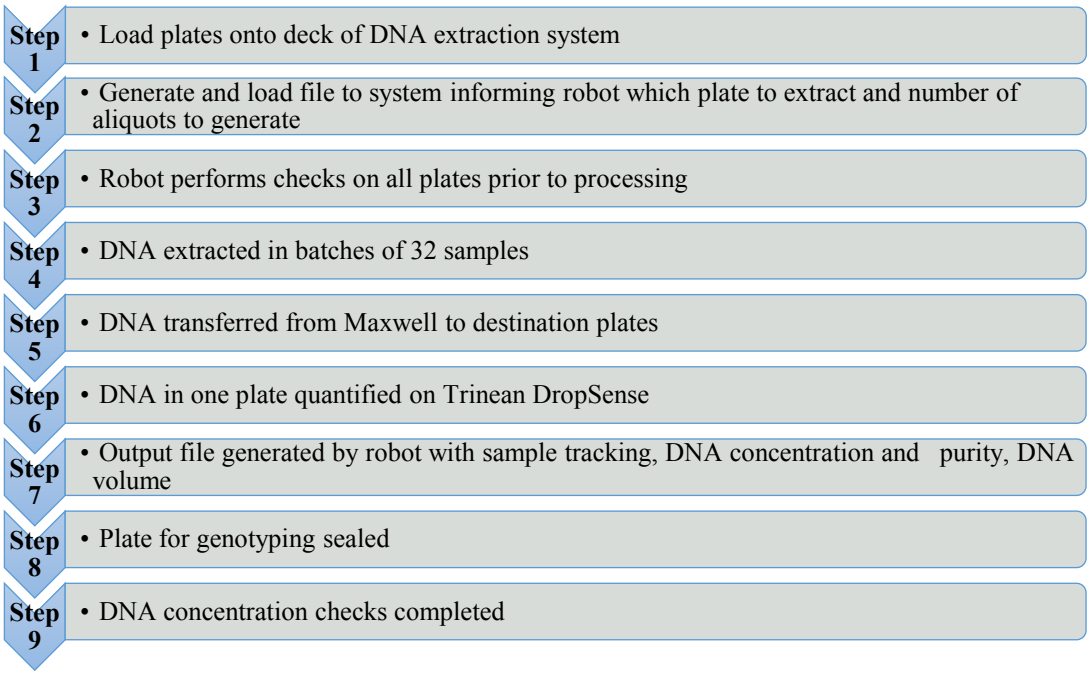


Figure 4 - 3: Overview of the automated process for DNA extraction.
(Source: adapted from (417)).

4.1.3.2 Genotyping and initial quality control (QC)

The Affymetrix research service laboratory (Santa Clara, California, USA) was responsible for genotyping the samples, generating the genotypic data and performing the initial quality control checks (417, 418). Extracted DNA samples were genotyped on two arrays, the UK BiLEVE array and UK Biobank Axiom Array. The UK BiLEVE array was applied on an initial ~50,000 (11 batches) individuals. UK Biobank Axiom Array was used to genotype the remaining ~450,000 (95 batches) samples. The two genotyping arrays were very similar to each other (common marker content >95%) and included a comprehensive coverage of genome-wide common and low frequency variants, rare coding variants, and genetic markers of specific interest (e.g., pharmacogenomic markers, human leukocyte antigen [HLA], inflammation, and expression quantitative trait loci [eQTL] variants). A summary of the marker content on the UK Biobank Axiom Array is shown in **Figure 4-4**. The positions of markers were reported based on the Genome Reference Consortium Human Reference 37 (GRCh37). Further details about the genotyping process and the design of UK Biobank Axiom Array are available at (419).

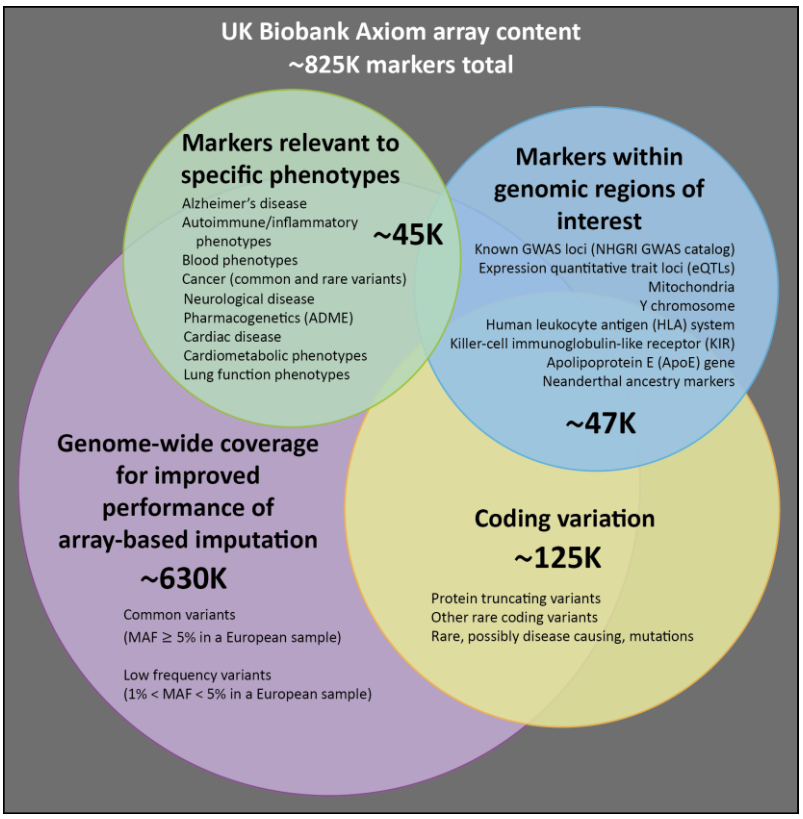


Figure 4 - 4: Summary of UK Biobank genotyping array content.

This is a schematic representation of the different categories of content on the UK Biobank Axiom array (Source: adapted from (419)).

An initial round of quality control (QC) of genotyping was conducted by the Affymetrix laboratory to exclude markers with poor cluster properties (418). In summary, variants which failed the basic Affymetrix genotyping quality metrics indicating poor genotype clustering (cluster QC) were excluded. This included the exclusion of variants for which (i) more than three genotype clusters were observed (indicating an off-target measurement), (ii) the call rate was less than 95%, or (iii) there was failure in one of the cluster quality metrics of Fisher's linear discriminant (FLD), Heterozygous cluster strength offset (HetSO), Homozygote Ratio Offset (HomRO) (thresholds were defined in the Affymetrix Axiom Genotyping Solution Data Analysis Guide). SNPs failing to meet the cluster QC Metrics were set to missing for all samples in that batch. More details about the Affymetrix calling algorithms and filtering protocols were documented in (418). The cluster QC conducted by Affymetrix resulted in a data set of 489,212 individuals typed at 812,428 unique markers proceeded to post-genotyping QC.

4.1.3.3 Additional post-genotyping QC

The Wellcome Trust Centre for Human Genetics (WTCHG) performed an additional round of marker-based QC to account for the population structure and check the consistency of genotype frequency between batches/plates. Details of the process of QC performed by the WTCHG is available at (420, 421). Briefly, to account for the population structure, they computed the SNP QC metrics by using a homogenous subset of European participants in UK Biobank. To identify homogenous individuals they projected UK Biobank samples on major principal components (PCs) computed by using the HapMap3 reference panel. Samples projected to the North-West European ancestry cluster were selected for computing the marker-based QC metrics. To detect any batch effects they tested whether the given batch had the same genotype frequencies as all other batches combined. Similarly, to look for any plate effects, they tested whether the given plate had the same genotype frequencies as all other plates, within the same batch. They also performed an exact Hardy-Weinberg Equilibrium (HWE) test for each batch within a homogenous sample. Genotypes at the SNPs that failed any of these tests with a p-value of $<10^{-12}$ were set to missing in that batch. Two tests (array effect and discordance across control replicates) were performed for each marker across all batches. Any marker that did not pass one of these two tests was excluded from the dataset for all batches. After the marker-based QC the final release of genotype data contained 805,426 SNPs (>99% of the array content). Examples of markers failing these QC tests are displayed in **Figure 4-5**.

Sample-based QC was then performed by the WTCHG based on a set of 605,876 high quality markers (genotyped on both arrays, passed QC in all batches) to identify samples with poor quality genotype calls, control for population structure (genetic ancestry/ethnicity) and to find related individuals. The PCs were further computed to indicate the genetic ancestry and account for population structure in other sample-based QC metrics (such as heterozygosity). The results of the PC analysis were then applied to adjust the heterozygosity and refine the relatedness inference. The sex of each individual was inferred based on the relative intensity of markers on the Y and X chromosomes. Samples with high missingness/heterozygosity rates or sex mismatch were not removed from the data release, but instead referred by a list of variables to indicate the insufficient data quality. The relatedness of samples was also indicated by a number of variables for further assessment. A small number of samples (835 in total) that were identified as sample duplicates (as opposed to identical twins), were likely mishandled in the laboratory or were withdrawn from the project, were excluded prior to the data release.

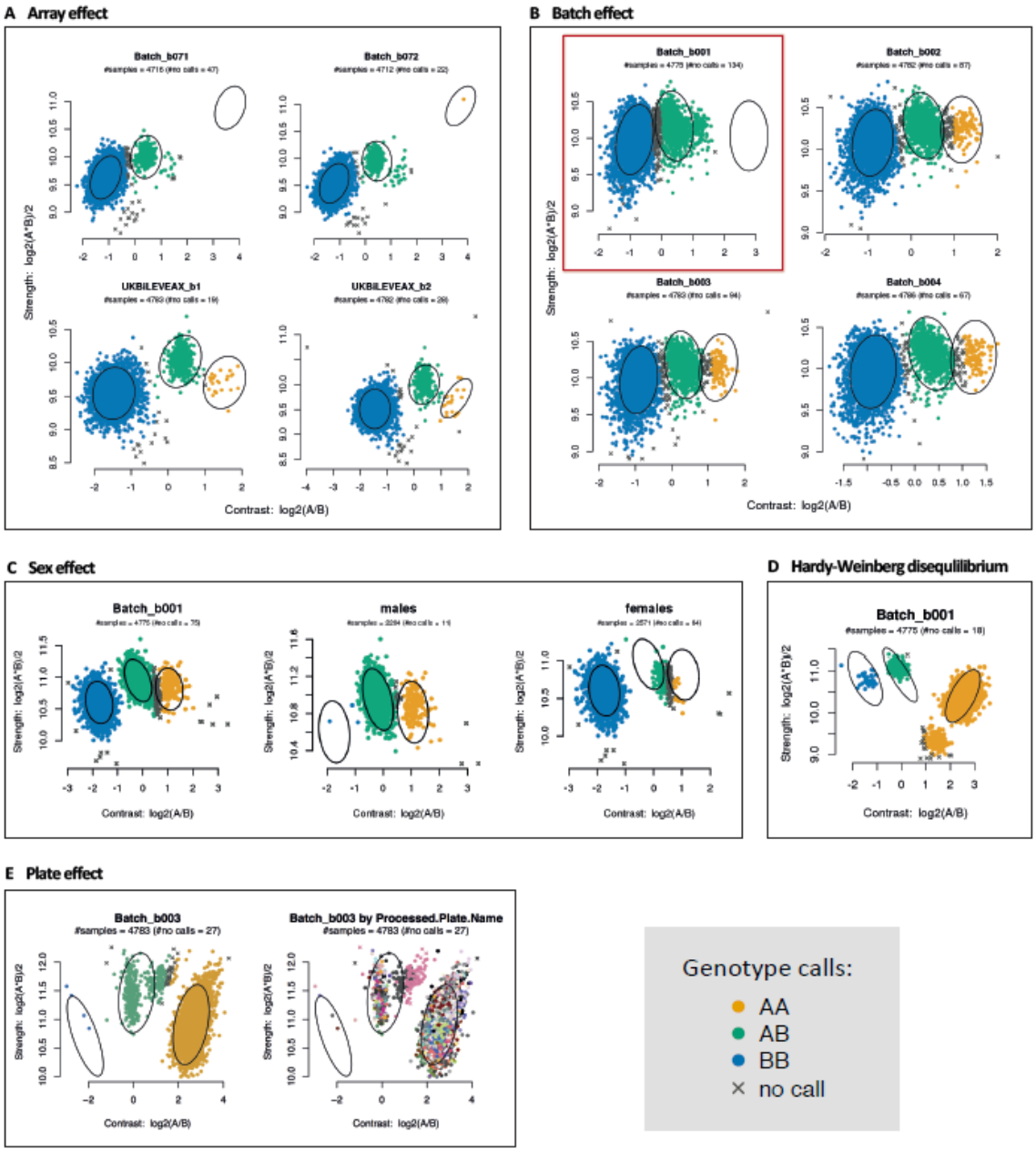


Figure 4 - 5: Examples of markers failing quality control tests.

Each sub-figure presents an example of markers failing the corresponding QC tests (Source: adapted from (421)).

4.1.3.4 Genotype imputation

The WTCHG was responsible for the genotype imputation. This process aimed to predict genotypes that were not directly assayed by the genotype arrays by using reference panels with a large number of haplotypes (422). The details of imputation for the UK Biobank

genotype data are presented in the reference (421). Briefly, the imputation process started with pre-phasing the genotyped markers and followed by a haploid imputation step.

- ***Pre-phasing***

A series of QC filters were applied first to select markers and samples for the phasing step. Genotyped markers were removed from the phasing step if they (i) were only present in either the UK BiLEVE or the UK Biobank Axiom arrays, (ii) failed the SNP QC in at least one of the batches, (iii) had a minor allele frequency < 0.0001, (iv) had >5% missingness. Samples were also removed if they were identified as outliers for heterozygosity and missingness. These filters resulted in a dataset of 670,739 autosomal SNPs in 487,442 samples. Phasing and imputation were carried out on the filtered dataset.

Phasing on the autosomes, in which a statistical method is applied to infer the underlying haplotypes of each individual, was carried out using the *SHAPEIT3* (423). The accuracy of the pre-phasing method was assessed by taking advantage of the 696 mother-father-child trios that were identified in the UK Biobank and the median switch error rate was estimated to be 0.229% (421).

- ***Imputation***

For the interim data release (~150,000), genotypes were imputed with a merged reference panel of the UK10K (424) and the 1000 Genomes Phase 3 (425), which consisted of ~ 87 million variants in 12,570 haplotypes. For the final release of the full UK Biobank data, genotypes were imputed by using the *IMPUTE4* (https://jmarchini.org/software/) with a combined reference panel of the Haplotype Reference Consortium (HRC) (426), UK10K haplotype resources (424), and the 1000 Genomes Phase 3 (425). This reference panel increased the number of testable markers to ~96 million variants and was expected to produce better imputation performance. In addition, they also imputed classical allelic variation at 11 human leukocyte antigen (HLA) genes localised in the histocompatibility complex (MHC) on chromosome 6, by using the HLA*IMP:02 algorithm with a multi-population reference panel (427). The imputation quality of the HLA genes were checked across the two imputation methods among samples of European ancestry and the reported call rates were >95.1% for all HLA loci (421).

The imputation process finally generated a dataset consisting of 92,693,895 autosomal SNPs for 487,442 individuals.

4.1.4 Medical records for longitude follow-up

A variety of sources and systems were used to follow up the disease occurrence, death and other health outcomes among enrolled participants (428). Currently, there were three different types of health records (i.e., hospital inpatient episodes, cancer registry data and death registry data) that have been incorporated into the central database. The process of incorporating routine electronic health records into the UK Biobank resource involved several steps (**Figure 4-6**).

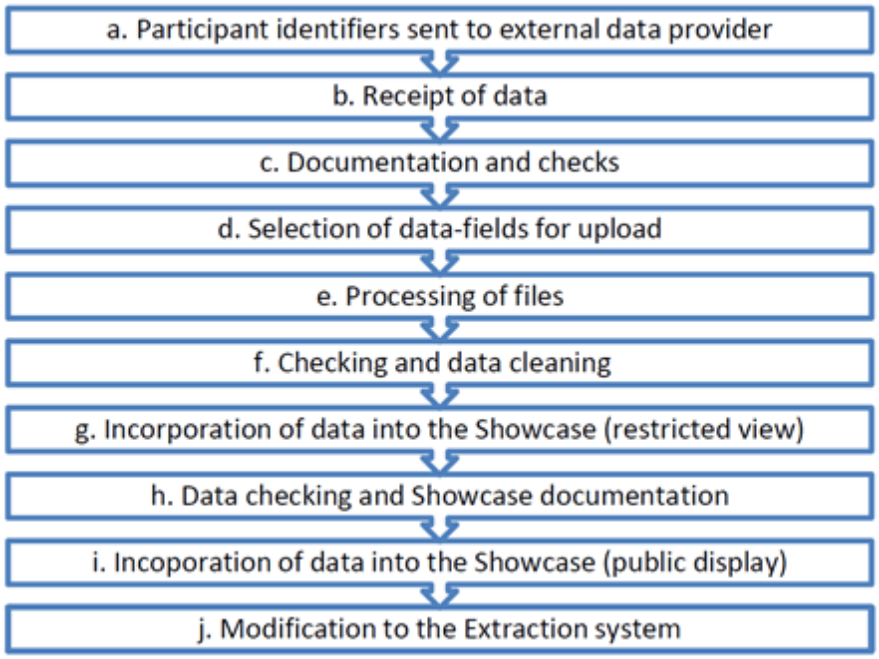


Figure 4 - 6: Steps involved in incorporating externally linked data into UK Biobank.
(Source: adapted from (428)).

4.1.4.1 Hospital inpatient data

UK Biobank collected information on health events experienced by participants via in-patient hospital records. Hospital admission data of England and Wales were collected at national level from the Department of Health's Hospital Episode Statistics (HES) and were managed by the NHS Health and Social Care Information Centre (HSCIC) (429). For Scotland, Scottish Morbidity Records (SMR) have been routinely collecting admission data of all Scottish NHS hospitals since 1980; these data are managed by the Information Services Division (ISD) of the NHS Common Service Agency. Historical hospital data back to 1996 in England & Wales and back to 1981 in Scotland have been provided to the UK Biobank as supplementary information prior to enrolment. Hospital inpatient data documented in the UK Biobank currently comprise five main domains:

- Admissions and discharge
- Diagnostic and operation codes (ICD-9/10)
- Maternity records
- Psychiatric census
- Critical care data

The availability of data types with corresponding dates is shown in **Table 4-4**. The diagnosis information of patients was recorded following the World Health Organisation's ICD and the operative procedures information were recorded following the OPCS (Office of Population, Censuses and Surveys: Classification of Interventions and Procedures).

Table 4 - 4：Source of hospital inpatient data.

Hospital admissions (in-patients)	**Country**	**Data provider**	**International classification of diseases (ICD)**		**Classification of interventions and procedures**		**Period of data currently available**
			ICD 9	ICD 10	OPCS3	OPCS4	
Hospital Episode Statistics (HES)	England	Heath & Social Care Information Centre (HSCIC)	--	1996-present	--	1996-present	1996-present
Patient Episode Database (PEDW)	Wales	Secure Anonymised Information Linkage (SAIL)	--	1999-present	--	1999-present	1999-present
Scottish Morbidity Record (SMR)	Scotland	Information and Statistics Division (ISD)	1981-1996	1996-present	1977-1988	1989-present	1981-present

(Source: adapted from (429)).

4.1.4.2 Cancer registry data

The cancer registration incorporated information on cancer diagnosis from a variety of sources including hospitals, cancer and treatment centres, hospices and nursing homes, private hospitals, cancer screening programmes, other cancer registries, general practices, death certificates, HES and Cancer Waiting Time (CWT) data (430). Cancer records were provided to UK Biobank by the Medical Research Information Service of the National

Health Service Information Centre (HSCIC) for participants from England and Wales and by the Information Services Division (ISD) of NHS Scotland for participants from Scotland. UK Biobank received details of cancer registration both prior to the inception of study (cancer registry data were available back to the early 1970s when the cancer registry was first established) and following the establishment of UK Biobank. The cancer registry data present in UK Biobank comprise (**Table 4-5**):

- Date of cancer diagnosis
- Age at cancer diagnosis
- Type of cancer (ICD-9/10)
- Reported occurrences of cancer
- Histology and behaviour code (ICD10-O-3)

The type of cancer was coded by either the ICD-9 or ICD-10 according to the time of registration. The histology and behaviour codes of neoplasms are presented as five-digit codes in ICD10-O-3, with the first four digits coding the histology and the fifth digit coding the behaviour.

Table 4 - 5: Source of cancer registry data.

Cancer registry	**Data provider**	**International classification of diseases (ICD)**		**Period of data currently available**
		ICD-9	ICD-10	
England & Wales	Heath & Social Care Information Centre (HSCIC)	1979-1994	1995-present	1996-present
Scotland	National Records of Scotland, NHS Central Register	1980-1996	1997-present	1981-present

Source: adapted from reference (430).

4.1.4.3 Death registry data

The follow-up of deaths was initiated at the recruitment phase of UK Biobank. UK Biobank participants were flagged at the NHS Central Registry from the date of their recruitment and UK Biobank received death notifications of participants (431). Data from the death certificates (all deaths and their certified causes) are sent to UK Biobank on a quarterly basis. The data presented in UK Biobank comprise (**Table 4-6**):

- Date of death
- Age at death
- Underlying (primary) cause of death: ICD-10

- Contributory (secondary) causes of death: ICD-10
- Description of cause of death

Table 4 - 6: Source of death registry data.

Death registry	Data provider	International classification of diseases (ICD)		Period of data currently available
		ICD-9	ICD-10	
England & Wales	Heath & Social Care Information Centre (HSCIC)	--	2006-present	April 2006-present
Scotland	Information and Statistics Division (ISD), Scotland	--	2006-present	April 2006-present

Source: adapted from reference (431).

4.2 Characteristics of the UK Biobank database

4.2.1 Summary of baseline data in UK Biobank

The UK Biobank includes 502,656 participants, of which 89% are from England, 7% from Scotland and 4% from Wales. The cohort consists of 94.1% (self-reported) white population (88.1% British, 2.6% Irish and 3.4% other white background) and 5.5% other ethnicities (2.28% Asian, 1.16% Black, 0.59% mixed and 1.46% other/unknown). There are 229,221 (45.6%) men and 273,445 (54.4%) women. The mean age of participants at recruitment was 56.53 years old (range: 40-69, SD: 8.10), with 23.8% aged at 40-49 years old, 33.6% aged at 50-59 years old and 42.6% aged at 60-69 years old. The mean BMI (body mass index) of participants at recruitment was 27.39 (range: 12.12-74.68, SD: 4.79) kg/m^2. According to the indices of multiple deprivation measured based on income, employment, health, education, access to services, community safety and physical environment, 48.2% of participants were characterised as low level of deprivation, 33.2% were characterised as moderate level of deprivation and 18.4% were characterised as high level of deprivation. The sociodemographic characteristics of UK Biobank participants are summarised in the **Table 4-7**.

Table 4 - 7: A summary of baseline characteristics of UK Biobank participants.

Continuous	**Mean ± S.D.**	**Number of participants**	**Representation (%)**
Age (years**)**	56.53 ± 8.10	502,620	99.9
Standing height (cm**)**	168.48 ± 9.27	500,130	99.5
Weight (kg**)**	77.97 ± 15.92	499,904	99.5
BMI (kg/m^2**)**	27.39 ± 4.79	499,579	99.4
DBP (mmHg**)***	82.02 ± 10.50	473,465	94.2
SBP (mmHg**)***	137.80 ± 19.27	473,460	94.2
Categorical	**Levels**	**Number of participants**	**Representation (%)**
Sex	Male	229,211	45.6
	Female	273,445	54.4
Age group	40-49 years	119,632	23.8
	50-59 years	168,892	33.6
	60-69 years	214,131	42.6
Self-reported ethnic background	White (British, Irish, and other white background)	472,798	94.06
	Asian (Indian, Pakistani, Chinese,	11,461	2.28

	Bangladeshi and other Asian background)		
	Black (Africa, Caribbean and other black background)	8,093	1.61
	Mixed (White & Asian, White & Black, other mixed background)	2,966	0.59
	Other/Unknown	7,339	1.46
Resident location	England	447,364	89.0
	Wales	35,186	7.0
	Scotland	20,106	4.0
Deprivation	Low	243,286	48.4
	Moderate	166,882	33.2
	High	92,489	18.4

*Blood pressures (Data-field *4079* [DBP] and Data-field *4080* [SBP]) were measured at the baseline assessment by using standard automated device (automated reading, two measures of blood pressure were taken a few moments apart).

4.2.2 Summary of health-outcome data in UK Biobank

As described above, UK Biobank incorporated three types of health records into the central database, including hospital inpatient records, cancer registry data and death registry data.

4.2.2.1 Hospital inpatient data

- ***Hospital inpatient episodes**:* A total of 361,234 hospital episodes were available in the UK Biobank database in 2016, covering 395,978 participants. In particular, there were 346,379 general episodes, 14,599 delivery episodes, 95 birth episodes, 94 psychiatric episodes and 50 other delivery/birth events (**Figure 4-7**).

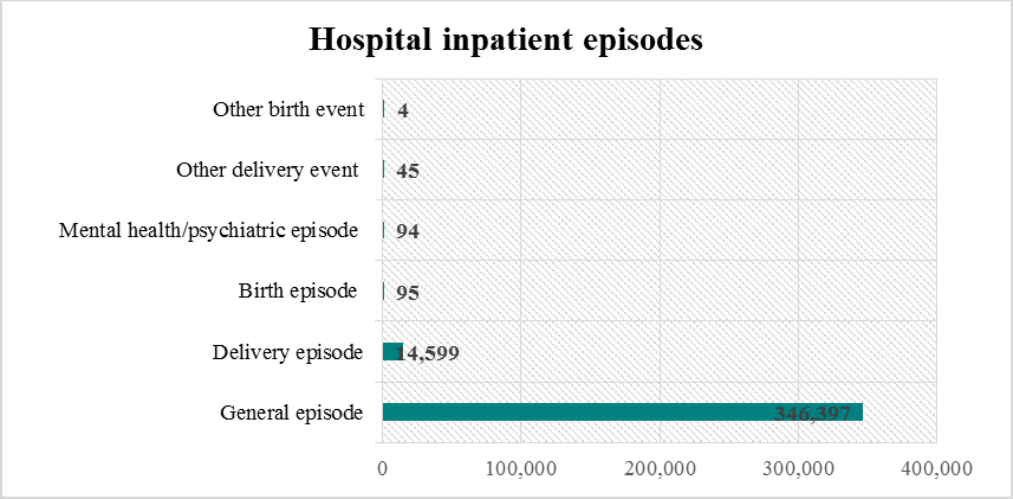


Figure 4 - 7: A summary of hospital episodes available in the UK Biobank database.

- ***ICD-9/10 codes for disease diagnosis:*** In the hospital inpatient data, each episode had a primary ICD-10 or ICD-9 diagnosis code to describe the event of hospitalisation, when applicable, one or more secondary ICD-10 or ICD-9 diagnosis codes follows to annotate the corresponding hospitalisation event. Thus, 361,234 hospital episodes recorded 1,491,042 items of primary ICD-10 code (8,160 distinct ICD-10 values) for 392,294 participants, 2,066,751 items of secondary ICD-10 code (10,528 distinct ICD-10 values) for 320,450 participants, 42,837 items of primary ICD-9 code (2,448 distinct ICD-9 values) for 20,309 participants and 17,293 items of secondary ICD-9 code (2,276 distinct ICD-9 values) for 8,716 participants. In summary, the hospital inpatient data included 10,528 unique ICD-10 codes covering 392,338 participants, and 2,449 unique ICD-9 codes covering 20,311 participants (**Table 4-8**).

Table 4 - 8: A summary of hospital inpatient data.

Diagnosis code	Data items	Distinct values	No. of participants
Main ICD-10 code	1,491,042	8,160	392,294
Secondary ICD-10 code	2,066,751	10,528	320,450
Sub-total	3,557,793	10,528	392,338
Main ICD-9 code	42,837	2,448	20,309
Secondary ICD-9 code	17,293	2,276	8,716
Sub-total	60,130	2,449	20,311
Total	**--**	**--**	**395, 978**

- ***Common prevalent diseases in hospital inpatient data:*** The individual ICD-10 codes were classified into 22 disease categories according to the ICD system: the number of cases included in each disease category is summarised in **Table 4-9** (range: 27-503,234; median: 50,218). Within the hospital episode data, the top 20 common prevalent ICD-10 codes (with the first 3 digits) and the corresponding diseases were identified and summarised in **Table 4-10**.

Table 4 - 9: A summary of the disease categories included in hospital inpatient data.

Disease category	Description	No. of cases
Chapter I	Certain infectious and parasitic diseases	35,981
Chapter II	Neoplasms	66,299
Chapter III	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	28,450
Chapter IV	Endocrine, nutritional and metabolic diseases	137,440
Chapter V	Mental and behavioural disorders	51,770
Chapter VI	Diseases of the nervous system	34,496
Chapter VII	Diseases of the eye and adnexa	27,258
Chapter VIII	Diseases of the ear and mastoid process	7,699
Chapter IX	Diseases of the circulatory system	257,515
Chapter X	Diseases of the respiratory system	86,615
Chapter XI	Diseases of the digestive system	202,376
Chapter XII	Diseases of the skin and subcutaneous tissue	20,844
Chapter XIII	Diseases of the musculoskeletal system and connective tissue	149,703
Chapter XIV	Diseases of the genitourinary system	111,762
Chapter XV	Pregnancy, childbirth and the puerperium	20,489
Chapter XVI	Certain conditions originating in the perinatal period	27
Chapter XVII	Congenital malformations, deformations and chromosomal abnormalities	5,162
Chapter XVIII	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	165,904
Chapter XIX	Injury, poisoning and certain other consequences of external causes	48,666
Chapter XX	External causes of morbidity and mortality	103,511
Chapter XXI	Factors influencing health status and contact with health services	503,234
Other	Codes for special purposes	1,550

Table 4 - 10: The top 20 prevalent ICD-10 codes/diseases in hospital inpatient data.

ICD-10	Disease description	No. of cases
I10	Essential (primary) hypertension	95,606
E78	Disorders of lipoprotein metabolism and other lipidaemias	44,949
R07	Pain in throat and chest	39,856
R10	Abdominal and pelvic pain	38,914
K29	Gastritis and duodenitis	35,786
K44	Diaphragmatic hernia	32,580
J45	Asthma	32,371
K57	Diverticular disease of intestine	32,304
I25	Chronic ischaemic heart disease	31,136
I84	Haemorrhoids	29,176
K62	Other diseases of anus and rectum	28,525
K21	Gastro-oesophageal reflux disease	28,100
M17	Gonarthrosis [arthrosis of knee]	24,283
E11	Type 2 diabetes mellitus	24,147
N39	Other disorders of urinary system	23,571
I20	Angina pectoris	22,422
Y83	Surgical operation and other surgical procedures as the cause of abnormal reaction	22,055
M19	Other arthrosis	19,966
K52	Other non-infective gastroenteritis and colitis	19,815
R69	Unknown and unspecified causes of morbidity	19,655

4.2.2.2 Cancer registry data

- ***Data items****:* A total of 233,875 data items, covering 79,111 participants, were available for cancer diagnosis, of which 76,321 data items were derived from the NHS information centre for Cancer Registry (2012 onwards), 8,468 data items were derived from the Scottish Cancer Registry (2012 onwards), and 63,002 data items were derived from the National Cancer Intelligence Network Cancer Registry (**Figure 4-8**).

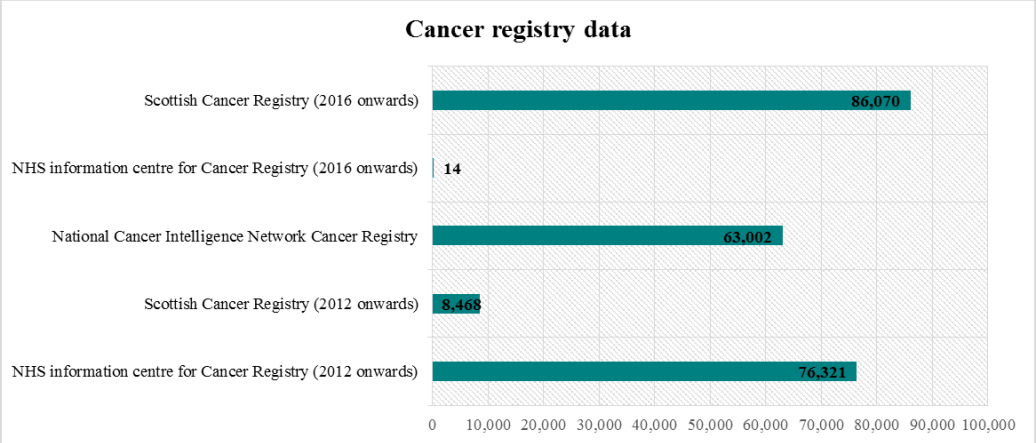


Figure 4 - 8: A summary of cancer registry data available in the UK Biobank database.

- ***ICD-9/10 for cancer diagnosis:*** Within the cancer registry data, 207,935 items were coded by ICD-10 system covering 73,841 participants, and 25,818 items were coded by ICD-9 system covering 11,227 participants. Overall, the database included 556 unique ICD-10 codes and 302 unique ICD-9 codes, covering 79,066 unique participants diagnosed with cancer **(Table 4-11)**.

Table 4 - 11: A summary of cancer registry data.

Disease classification	Data items	Distinct values	No. of participants
Type of cancer: ICD-10	207,935	556	73,841
Type of cancer: ICD-9	25,818	302	11,227
Total	**233,753**	**858**	**79,066**

- ***Common prevalent cancers in cancer registry data:*** The types of cancer coded by ICD-10 codes were classified according to the sites of neoplasms. The number of cases included in each sites of neoplasms were summarised in **Table 4-12** (range: 1-19,620; median: 2,032). Within the cancer registry data, the top 20 common prevalent cancers (with the first 3 digits of ICD-10 code) were identified and summarised in **Table 4-13**.

Table 4 - 12: A summary of the types of cancer according to the sites of neoplasms.

Sites of neoplasms	No. of cases*
Malignant neoplasms of lip, oral cavity and pharynx	913
Malignant neoplasms of digestive organs	6,563
Malignant neoplasms of respiratory and intrathoracic organs	1,978
Malignant neoplasms of bone and articular cartilage	83
Melanoma and other malignant neoplasms of skin	19,620
Malignant neoplasms of mesothelial and soft tissue	603
Malignant neoplasm of breast	13,025
Malignant neoplasms of female genital organs	3,748
Malignant neoplasms of male genital organs	7,942
Malignant neoplasms of urinary tract	2,087
Malignant neoplasms of eye, brain and other parts of central nervous system	615
Malignant neoplasms of thyroid and other endocrine glands	630
Malignant neoplasms of ill-defined, secondary and unspecified sites	468
Malignant neoplasms, stated or presumed to be primary, of lymphoid, hematopoietic and related tissue	3,738
Malignant neoplasms of independent (primary) multiple sites	1
In situ neoplasms	11,218
Benign neoplasms	1,009
Neoplasms of uncertain or unknown behavior	2,453

*Including both incident and prevalent cases.

Table 4 - 13: The top 20 common prevalent cancers in cancer registry data.

ICD-10	Type of cancer	No. of cases*
C44	Other malignant neoplasms of skin	18,628
C50	Malignant neoplasm of breast	13,663
C61	Malignant neoplasm of prostate	8,379
D06	Carcinoma in situ of cervix uteri	3,972
C18	Malignant neoplasm of colon	3,358
C43	Malignant melanoma of skin	3,290
D05	Carcinoma in situ of breast	2,455
C34	Malignant neoplasm of bronchus and lung	2,301
C54	Malignant neoplasm of corpus uteri	1,697
C20	Malignant neoplasm of rectum	1,562
D04	Carcinoma in situ of skin	1,446
D03	Melanoma in situ	1,276
C67	Hereditary factor IX deficiency	1,196
C64	Malignant neoplasm of bladder	1,187
C56	Malignant neoplasm of ovary	1,177
C83	Non-follicular lymphoma	961
D41	Neoplasm of uncertain or unknown behaviour of urinary organs	912
D09	Carcinoma in situ of other and unspecified sites	717
C91	Lymphoid leukaemia	683
C15	Malignant neoplasm of oesophagus	645

*Including both incident and prevalent cases.

4.2.2.3 Death registry data

- ***Data items:*** A total of 16,505 data items, covering 14,423 participants, were available from the death registry data (**Figure 4-9**), of which 13,217 data items were derived from the NHS information centre (10,252 items from post-2012 and 2,965 items from pre-2012), and 3,288 data items were derived from the Scottish Morbidity Record (726 items from post-2015 and 2,562 items from pre-2015).

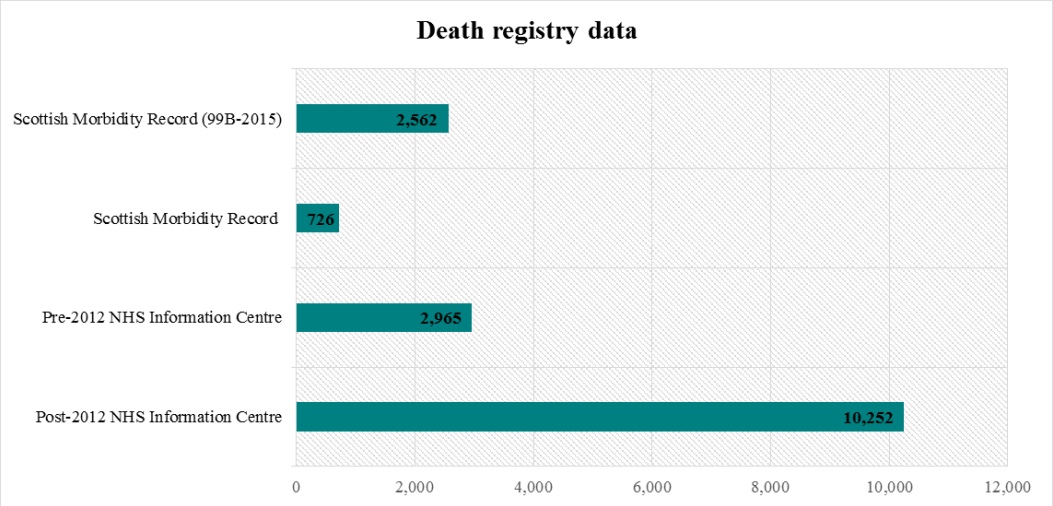


Figure 4 - 9: A summary of death registry data available in the UK Biobank database.

- ***ICD-10 codes for cause of death:*** Within the death registry data, a total of 37,293 data items were available to record the causes of death of 14,418 participants. Primary (underlying) causes of death (16,500 data items) were reported for all 14,418 participants, consisting of 888 unique ICD-10 codes. Contributory/secondary causes of death (20,793 data items) were reported for 8,782 participants, consisting of 1,144 unique ICD-10 codes. Overall, the database included 1,525 unique ICD-10 codes to annotate the causes of death for 14,418 participants (**Table 4-14**).

Table 4 - 14: A summary of death registry data.

Cause of death	Data items	Distinct values	No. of participants
Primary (underlying) cause of death: ICD10	16,500	888	14,418
Secondary (contributory) cause of death: ICD10	20,793	1,144	8,782
Total	**37,293**	**1,525**	**14,418**

- ***Common causes of death in death registry data:*** The primary and secondary causes of death coded by ICD-10 codes were classified into 22 disease categories according to the ICD system: the number of deaths caused by each disease category were summarized in **Table 4-15** (range: 27-503,234; median: 50,218). Within the death registry data the top 20 common prevalent ICD-10 codes (with the first 3 digits) and the corresponding causes of death were identified and summarised in **Table 4-16**.

Table 4 - 15: A summary of the primary and secondary death causes in each disease category.

Disease category	Description	Primary causes of death	Secondary causes of death
Chapter I	Certain infectious and parasitic diseases	146	791
Chapter II	Neoplasms	9,577	2,905
Chapter III	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	33	193
Chapter IV	Endocrine, nutritional and metabolic diseases	153	1,174
Chapter V	Mental and behavioural disorders	129	338
Chapter VI	Diseases of the nervous system	517	503
Chapter VII	Diseases of the eye and adnexa	0	2
Chapter VIII	Diseases of the ear and mastoid process	0	2
Chapter IX	Diseases of the circulatory system	3,438	6,248
Chapter X	Diseases of the respiratory system	943	3,238
Chapter XI	Diseases of the digestive system	669	1,268
Chapter XII	Diseases of the skin and subcutaneous tissue	14	56
Chapter XIII	Diseases of the musculoskeletal system and connective tissue	90	227
Chapter XIV	Diseases of the genitourinary system	84	801
Chapter XV	Pregnancy, childbirth and the puerperium	34	0
Chapter XVI	Certain conditions originating in the perinatal period	0	0
Chapter XVII	Congenital malformations, deformations and chromosomal abnormalities	82	33
Chapter XVIII	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	576	738
Chapter XIX	Injury, poisoning and certain other consequences of external causes	0	1,142
Chapter XX	External causes of morbidity and mortality	576	1,121
Chapter XXI	Factors influencing health status and contact with health services	0	10
Other	Codes for special purposes	14	0

Table 4 - 16: The top 20 common prevalent death causes in death registry data.

ICD-10	Disease description	No. of cases*
C34	Lung cancer	1,786
I25	Chronic ischaemic heart disease	1,089
C50	Malignant neoplasm of breast	841
I21	Acute myocardial infarction	819
C25	Malignant neoplasm of pancreas	617
C18	Malignant neoplasm of colon	526
C61	Malignant neoplasm of prostate	485
C71	Malignant neoplasm of brain	468
C15	Malignant neoplasm of oesophagus	434
C80	Malignant neoplasm, without specification of site	414
J44	Other chronic obstructive pulmonary disease	341
C56	Malignant neoplasm of ovary	340
C22	Malignant neoplasm of liver and intrahepatic bile ducts	297
C64	Malignant neoplasm of kidney, except renal pelvis	265
C16	Malignant neoplasm of stomach	242
C45	Mesothelioma	235
J84	Other interstitial pulmonary diseases	234
J18	Influenza and pneumonia	198
C19	Malignant neoplasm of rectosigmoid junction	195
C20	Malignant neoplasm of rectum	192

*Including both incident and prevalent cases.

4.2.3 Summary of genotype data in UK Biobank

4.2.3.1 Genotype data quality

The application of marker-based and sample-based QC processes resulted in a released genotypic dataset of 805,426 markers for 488,377 samples. The number and proportion of SNPs which failed marker-based quality tests are summarised in **Table 4-17**. Specifically, 197 SNPs (proportion: 0.25‰) failed the batch effect test, 284 SNPs (proportion: 0.36‰) failed the plate effect test, 572 SNPs (proportion: 0.72‰) violated the HWE test, 45 SNPs (proportion: 0.06‰) failed the sex effect test, 5,417 SNPs (proportion: 6.83‰) failed the array effect test and 45 SNPs (proportion: 0.06‰) were in discordance in genotyping calls against the controls. Overall, the proportion of SNPs removed because of poor genotyping call was 9.7‰.

Table 4 - 17: Failure rates in marker-based quality control.

QC steps	Average number of SNPs failed per batch (SD)	Proportion of all genotype calls affected (‰)
Marker-based QC		
Batch effect	197 (86)	0.25
Plate effect	284 (266)	0.36
Violation of HWE	572 (77)	0.72
Sex effect	45 (5)	0.06
Array effect	5,417	6.83
Discordance across control replicates	622	0.80
Total	**7,704 (721)**	**9.71**

A total of 3,163 genotyped samples (6.5‰) were identified as poor quality based on the sample-based QC (**Table 4-18**). After adjusting heterozygosity with the first six PCs of population structure, 968 samples (2.0‰) were identified as having an unusually high heterozygosity and/or >5% missing rate. When comparing the self-reported sex and the genetic inferred sex, 378 samples (0.8‰) were identified as a sex mismatch. Additionally, 652 samples (proportion: 1.3‰) were indicated as instances of sex chromosome aneuploidy. Another 188 samples (proportion: 0.4‰) that appeared to be related (3rd degree) to a very large number (>10) of individuals were suggested to be excluded, as the excess related pairs were likely to be false positives. Another set of 977 samples (proportion: 2.0‰) were indicated to be excluded by the kinship estimation, as they had properties (e.g., high missing rates) that would lead to unreliable kinship estimates.

Table 4 - 18: Failure rates in sample-based quality control.

Sample-based QC	Number of samples failed QC	Proportion of samples affected (‰)
High heterozygosity and/or >5% missing rate	968	2.0
Sex mismatch	378	0.8
Putative sex-chromosome aneuploidy	652	1.3
Excess related pairs	188	0.4
Unreliable kinship estimates	977	2.0
Total	**3,163**	**6.5**

4.2.3.2 Genetic population structure

The population structure of UK Biobank is diverse as indicated by the self-reported ethnic background (**Table 4-19**). Genetic PCs were calculated to quantify the population structure from genetic ancestry/ethnicity. The top 40 PCs were calculated using the *fastPCA* (432). Then the SNP-loads for each PC were computed by carrying out the appropriate matrix multiplications based on mean-centred and variance-scaled genotypes and the PC scores were computed by the *fastPCA*. All samples were then projected onto the PCs using their SNP-loads. As expected, individuals with similar PC scores have similar self-reported ethnic backgrounds. The overall population structure of UK Biobank as revealed by the first six pairs of PCs is shown in **Figure 4-10**. Of the 488,377 samples, 409,694 individuals (proportion: 83.9%) were identified as White with their PC scores falling in the neighbourhood of the North-West European ancestry cluster. This was in concordance with the self-reported ethnicity, where a majority of the UK Biobank participants reported their ethnic background as "British" (within the broader-level group "White"). When analysing the PCs within the self-reported British, a homogeneous population of White British ancestry subset (defined with a combination of self-reported ethnic background as "British" and genetic indicated ethnic background as "White" based on the PCs) was identified, which included 409,703 individuals.

Table 4 - 19: Self-reported ethnic background.

Self-reported ethnic background	No. of participants	Percentage (%)
White (British, Irish, and other white background)	472,798	94.06
Asian (Indian, Pakistani, Chinese, Bangladeshi and other Asian background)	11,461	2.28
Black (Africa, Caribbean and other black background)	8,093	1.61
Mixed (White & Asian, White & Black, other mixed background)	2,966	0.59
Other/Unknown	7,339	1.46

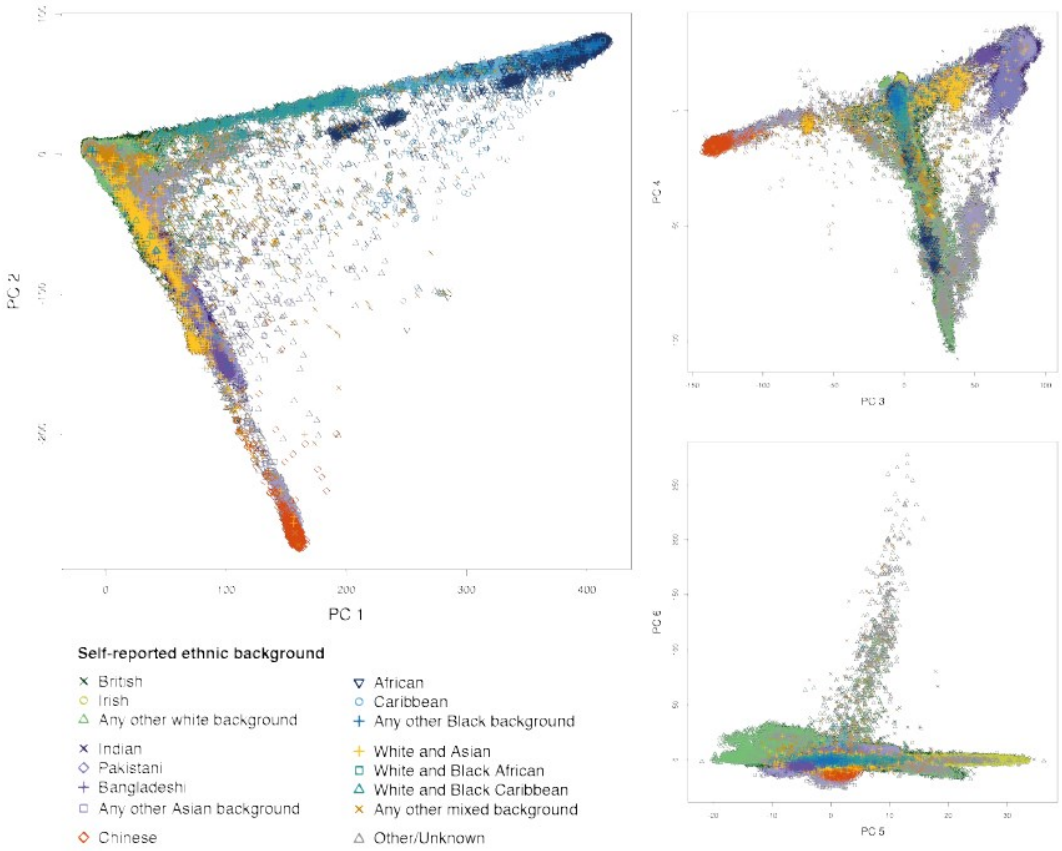


Figure 4 - 10: Ancestral diversity in the UK Biobank cohort.

Plots were made of consecutive pairs of the first six PCs in a PC analysis of genotype data for UK Biobank participants. Each point represents an individual and is placed according to their principal component scores (using genetic data only), with shapes and colours indicating their self-reported ethnic background as shown in the legend (Source: adapted from (417)).

4.2.3.3 Genetic relatedness

The genetic relatedness among UK Biobank participants was estimated by the kinship coefficients for all pairs of samples. According to the kinship inference, a total of 147,731 UK Biobank participants (30.3%) were inferred to be related (3rd degree or closer) to at least one other participant in the cohort. The related individuals formed a total of 107,162 related pairs (**Table 4-20**), including 179 pairs of monozygotic twins, 6,276 pairs of parent-offspring, 22,666 pairs of full siblings, 11,113 pairs of 2nd degree relatives (e.g., uncle-niece) and 66,928 pairs of 3rd degree relatives (e.g., first cousins). In order to find the subset with the maximum number of unrelated individuals, further examination of the relationships among the samples within family groups was required (the procedure of identifying the largest possible subset of unrelated individuals is described below in the *Chapter 4, Section 4.3.1 "Study population selection"*).

Table 4 - 20: Summary of related pairs (3rd degree or closer) for UK Biobank.*

	Monozygotic twins	Parent-offspring	Full siblings	2nd degree	3rd degree	Total
Number of pairs	179	6,276	22,666	11,113	66,928	107,162

*Counts are derived from the kinship coefficients.

4.3 Cleaning and preparation of UK Biobank data

4.3.1 Study population selection

The SUA PheWAS analysis was constrained to a subset of unrelated White British individuals with high quality genotype data in order to minimize the influence of diverse population structure within UK Biobank. The metrics used to select the target study population were based on the data fields created for the genotype QC in UK Biobank. The detailed process for selecting target population is described below.

4.3.1.1 Sex mismatch

There are two data fields available to describe the sex of samples. Field one is the self-reported sex submitted by participants (coded by ***data field 31*** as "male" and "female") and the other is the genetic sex (coded by ***data field 22001*** as "male" and "female") inferred from the calling genotypes on the male-specific region of the Y chromosome and the non-pseudoautosomal region of the X chromosome. When self-reported sex was not consistent with the inferred sex from genotype data samples were referred to as a sex mismatch. When comparing the ***data field 31*** with the ***data field 22001***, 99.9% samples showed concordance, but for a small number of samples (n=378) the data fields did not match and were thus excluded from the target population.

4.3.1.2 Outliers in heterozygosity and missing rates

The property of outliers in heterozygosity and high missing rates was coded as 0 (no) or 1 (yes) and was described in the variable: "*het.missing.outliers*". A total of 968 samples (coded as 1 ["yes"]) were identified as outliers and were excluded from the study population.

4.3.1.3 Putative aneuploidy in sex chromosome

The property of putative aneuploidy in sex chromosome (putatively carrying sex chromosome configurations that are not either XX or XY) was coded as 0 ("no") or 1 ("yes") and was described in the variable: "*putative.sex.chromosome.aneuploidy*". A total of 652 samples (coded as 1 ["yes"]) were identified as aneuploidy in sex chromosome and thus were excluded from the study population.

4.3.1.4 Excess relatives

The property of excess relatives (with more than 10 putative 3rd degree relatives in the kinship table) was coded as 0 ("no") or 1 ("yes") and was described in the variable:

"*excess.relatives*". A total of 188 samples (coded as 1 ["yes"]) were identified to have excess relatives in UK Biobank and thus excluded from the study population.

4.3.1.5 White British ancestry

The property of White British ancestry (self-reported ethnic background as "British" and genetic ethnic background as "White") was coded as 0 ("no") or 1 ("yes") and was described in the variable: "*in.white.British.ancestry.subset*". A total of 78,674 samples (coded as 0 ["no"]) did not belong to the White British subset and thus were excluded from the study population.

4.3.1.6 Subset of unrelated individuals

To find the subset with the maximum number of unrelated individuals, Bycroft *et al* developed a procedure by using the R package "*i-graph (v1.0.1)*". I first pruned the full pairwise kinship table so that it only included White British individuals that passed the sample quality control; I then converted the pruned kinship table into a graph object where each vertex is an individual and edges exist between pairs of related individuals (421). Then, for each "family" (i.e. a network of nodes joined by edges as shown in **Figure 4-11**), the largest subset of individuals (vertices) without relatedness were identified and chosen by an algorithm implemented in the "*largest_ivs*" function in *"i-graph"* R package. For instance, in a simple case of trios, the child would be excluded, leaving the maximum number of two unrelated parents. When there were multiple possible solutions for the choice of the largest subset of unrelated individuals, one of these solutions were chosen at random. By following this procedure, the largest possible subset was identified, it included 339,256 unrelated individuals when restricted to the quality-filtered subset of White British.

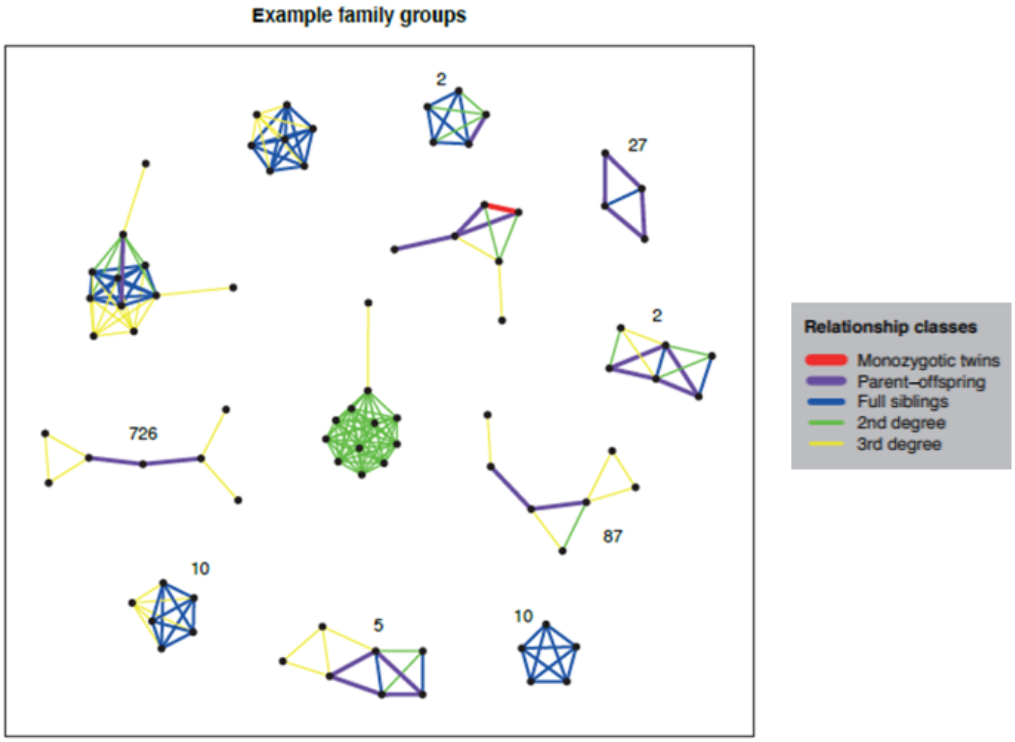


Figure 4 - 11: Familial relatedness in the UK Biobank cohort.

This figure illustrates examples of family groups within the UK Biobank cohort. Points indicate participants and lines between points indicate familial relatedness (3rd degree and closer), as inferred from the genetic data. The colour and thickness of the lines indicate different relative classes, as shown in the key. No integer means there is only the one shown; for example, there is only one network that comprises 6 full siblings (plus one 3rd degree relative who is related to all siblings). An integer next to a network indicates the total number of family networks in the cohort with the same configuration, ignoring 3rd degree pairs. For example, there are 10 networks that comprise exactly 5 full siblings (two examples, which differ with respect to a 3rd degree relative, are shown on this plot); when selecting the subset of unrelated individuals, one of the 5 full sibling were retained at random; (Source: adapted from (421)).

4.3.2 Covariate selection

Covariates for the SUA PheWAS analysis were selected based on consideration of the baseline characteristics of UK Biobank cohort, the properties of genotyping data and the potential confounding factors related to SUA level. The following list of potential covariates were included: age, sex, BMI, assessment centre, the first 6 PCs, deprivation index, smoking status and alcohol intake frequency. Some other potential confounding variables such as data on creatinine concentration, HDL cholesterol and triglycerides were not available in UK Biobank and thus were not included as covariates.

4.3.2.1 Sex, age and BMI

Among the selected study population (with 339,256 unrelated White British individuals), there were 182,110 females and 157,146 males (*data field 32*). To describe the age of study populatio, two data fields are available: age at recruitment (*data field 21022*) and age when attended assessment centre (*data field 21003*). The mean age of the study population at 2016 was 64.82 (SD: 8.00) years. The distribution of male and female participants in each age group is shown in **Figure 4-12**. The BMI (defined as: weight (kg) / [height (m)]^2) was coded by the *data field 21001* as a continuous variable expressed in kg/m^2. The mean value of BMI was 27.02 (SD: 4.76) kg/m^2 for female and 27.83 (SD: 4.23) kg/m^2 for male (**Figure 4-13**).

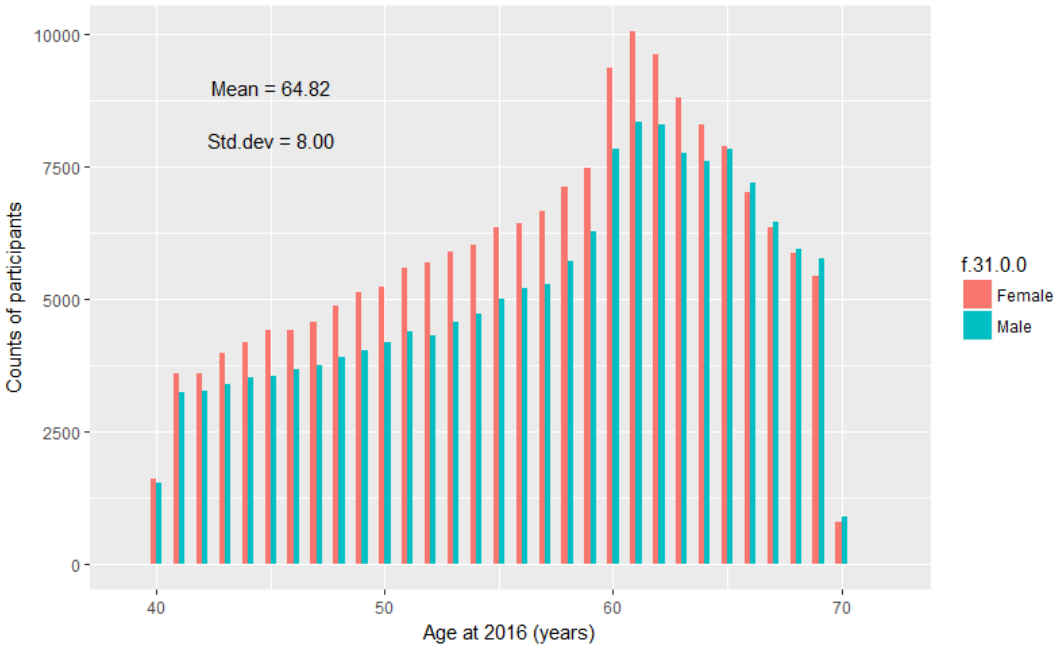


Figure 4 - 12: The age distribution of male and female participants.

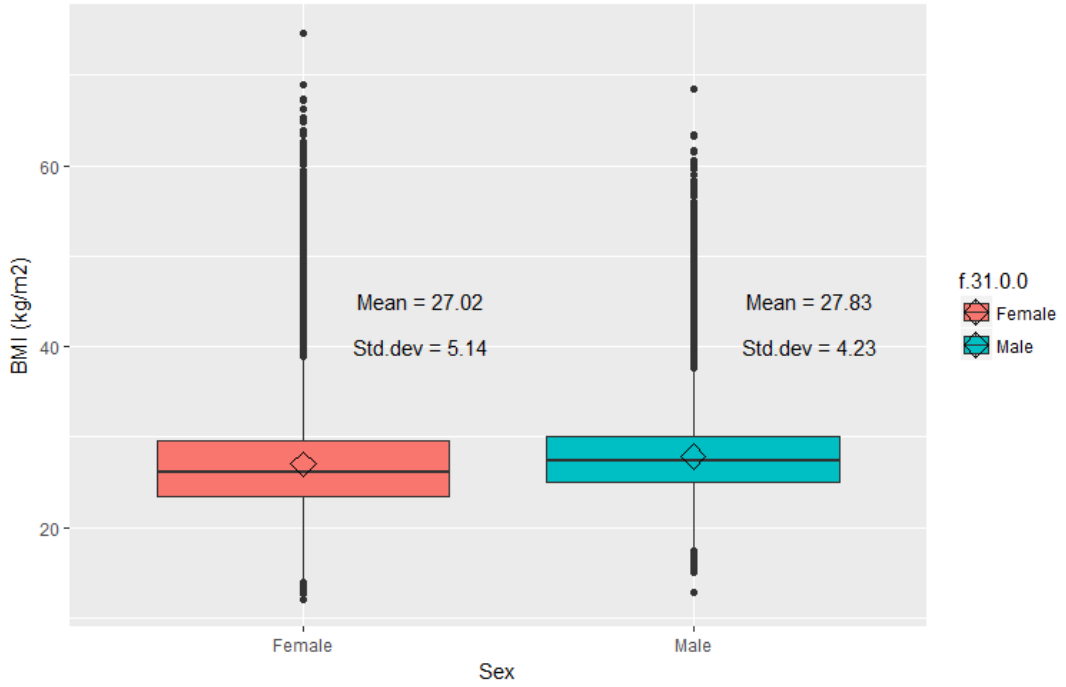


Figure 4 - 13: The boxplot for BMI (kg/m^2) in each sex.

4.3.2.2 Assessment centre

Assessment centre, where the participants were recruited, is coded by the *data field* ***f.54.0.0.*** The 339,256 unrelated White British individuals included in this study were distributed across 22 assessment centres. The number of participants recruited from each assessment centre is displayed in **Figure 4-14**. The assessment centre with the largest number of participants in this study was Leeds (n=31,348). The characteristics of the other covariates (sex, age, BMI) are summarised for each assessment centre and are shown in **Table 4-21**.

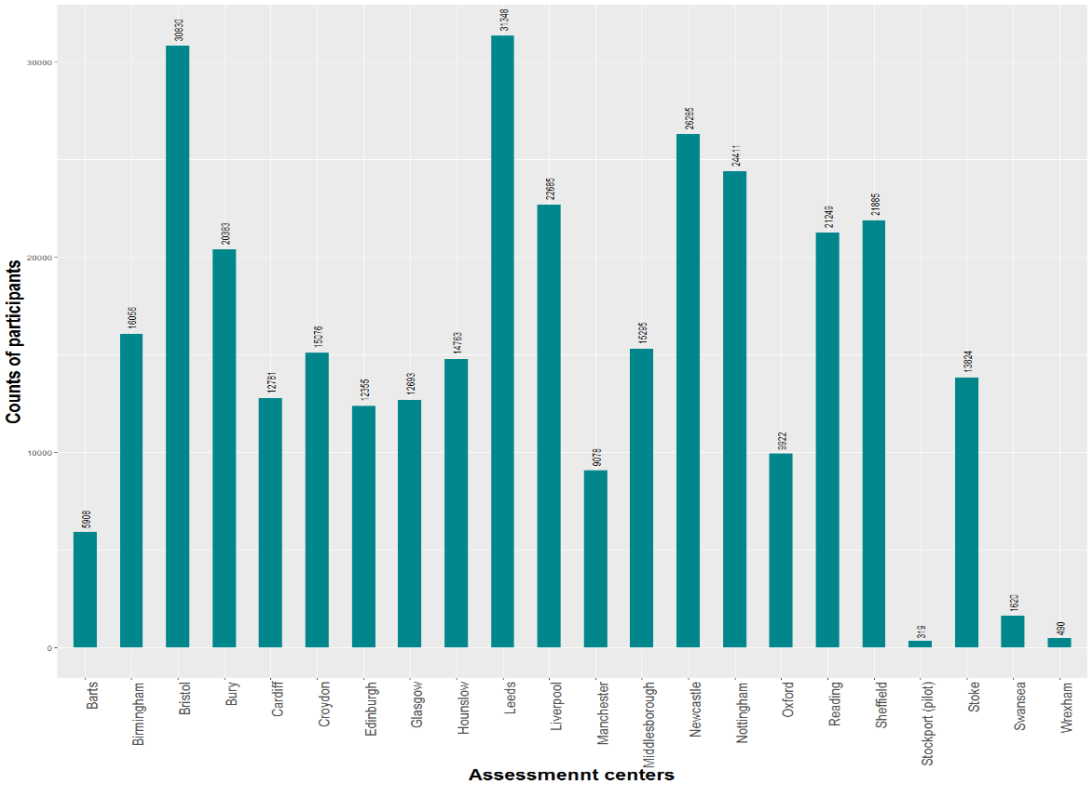


Figure 4 - 14: The number of participants recruited from each assessment centre.

Table 4 - 21 A summary of the baseline characteristics of participants recruited from each assessment centre.

Assessment centre	**Counts**	**Sex** (*female/male*)	**Age** (years)	**BMI** (kg/m^2)
Barts	5,908	3,103/2,805	63.47±8.13	26.24±4.79
Birmingham	16,056	8,385/7,671	63.92±8.11	27.70±4.85
Bristol	30,830	16,982/13,848	63.98±8.27	27.02±4.66
Bury	20,383	10,629/9,691	65.87±7.76	27.70±4.70
Cardiff	12,781	6,885/5,896	65.01±7.84	27.97±4.86
Croydon	15,076	8,217/6,859	64.28±7.79	26.98±4.70
Edinburgh	12,355	6,826/5,529	65.39±7.92	27.02±4.67
Glasgow	12,693	6,983/5,710	65.45±8.16	27.65±4.85
Hounslow	14,763	7,922/6,841	64.22±7.98	26.71±4.70
Leeds	31,348	16,984/14,364	64.97±7.92	27.44±4.72
Liverpool	22,685	12,105/10,580	64.64±7.9	27.83±4.90
Manchester	9,078	4,804/4,274	65.15±8.34	27.46±4.94
Middlesborough	15,295	8,103/7,192	64.32±8.06	27.81±4.77
Newcastle	26,285	14,184/12,101	65.2±7.94	27.73±4.74
Nottingham	24,411	13,129/11,282	65.14±7.92	27.38±4.70

Oxford	9,922	5,654/4,268	66.04±7.96	26.67±4.62
Reading	21,249	11,489/9,760	64.97±7.91	26.87±4.50
Sheffield	21,885	11,651/10,234	64.36±7.82	27.53±4.81
Stockport (pilot)	319	169/150	65.84±7.52	26.46±4.90
Stoke	13,824	6,703/7,121	65.45±8.00	27.75±4.69
Swansea	1,620	882/738	64.37±7.98	28.19±5.18
Wrexham	490	258/232	63.53±8.22	28.62±5.81
Overall	**339,256**	**182,110/157,146**	**64.82±8.00**	**27.40±4.76**

4.3.2.3 Genetic principal components

As explained in *Section 2.3.2*, the top 40 genetic PCs were computed and represented by a set of PC scores for all samples in the UK Biobank cohort. The top 40 PC scores were coded by the ***data field 22009*** with an array index (multiple data items for each instance) running from 1 to 40. The first six PC scores were chosen as covariates to adjust for the heterozygosity of the population structure in SUA PheWAS, as they explained the most variation. The plots for the first six PC scores (pairs of PCs: 1&2; 3&4; 5&6) for the UK Biobank samples are shown in **Figure 4-15** and the blue crosses represent the subset of White British ancestry (the target population of SUA PheWAS analysis) with their first six PC scores falling in the neighbourhood of the North-West Europe ancestry cluster.

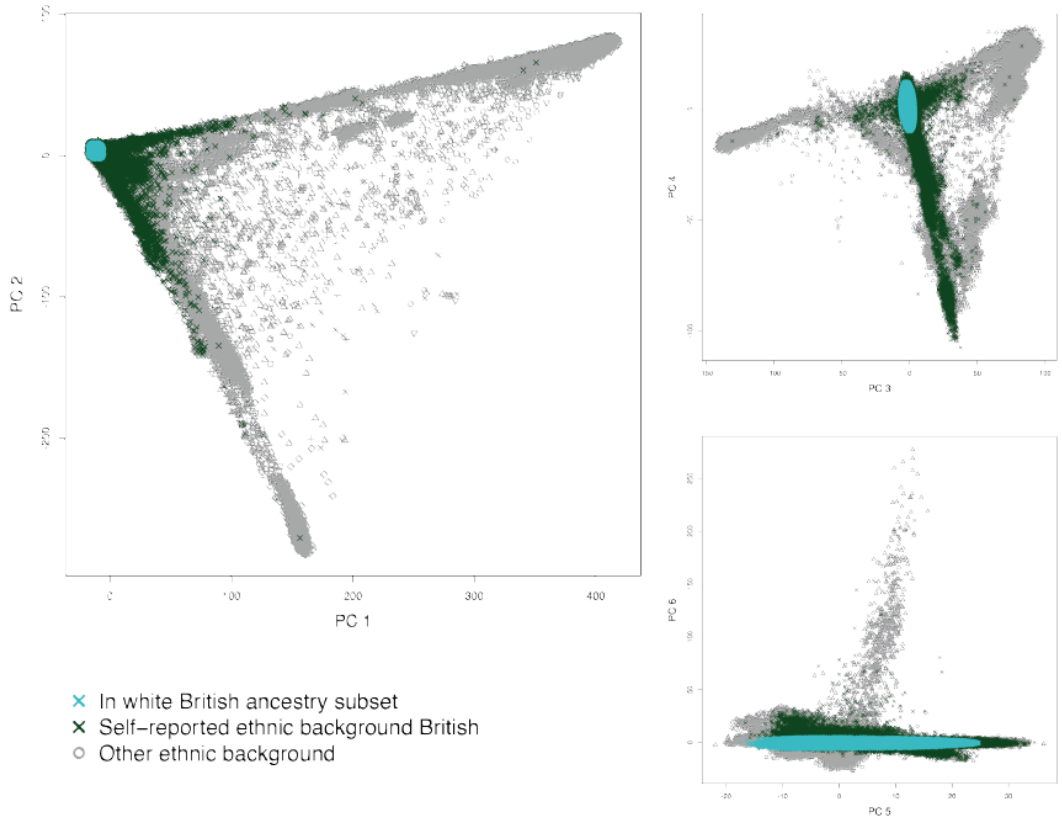


Figure 4 - 15: Plots for the first six PC scores.

Each plot shows the principal component scores for all UK Biobank samples used to select the white British ancestry subset. Non-grey points indicate participants who have self-reported ethnic background "British" and participants with other ethnic backgrounds are coloured grey (Source: adapted from reference (421)).

4.3.2.4 Other covariates

To adjust for any effect of environmental exposures, smoking status and alcohol intake frequency were also included as covariates. Smoking status (coded by ***data field 20116***) was categorised as "current smoking", "previous smoking", "never smoking" and "prefer not to answer". The number of participants assigned to each smoking status is shown in **Table 4-22**. Alcohol intake frequency (coded by ***data field 1558***) was grouped as "daily or almost daily", "three or four times a week", "once or twice a week", "one to three times a month", "special occasion only", "never", and "prefer not to answer". The number of participants grouped in each category is summarised in **Table 4 - 22**.

Table 4 - 22: The smoking status and alcohol intake frequency among participants.

Categorical	Levels	No. of participants	Proportion (%)
Smoking status	Current smoking	35,798	29.0
	Previous smoking	31,090	25.2
	Never smoking	55,376	44.9
	Prefer not to answer	1,029	0.8
	Missing	88	<0.1
Alcohol intake frequency	Daily or almost daily	25,653	21.4
	3 or 4 times a week	28,224	23.5
	1 or 2 times a week	31,434	26.2
	1 to 3 times a month	13,373	11.1
	Special occasions only	13,161	11.0
	Never	8,168	6.8
	Prefer not to answer	78	<0.1
	Missing	0	0

4.3.3 Construction of genetic instruments

4.3.3.1 Genetic risk loci and effect sizes

To generate genetic instruments for SUA levels I searched for genetic variants associated with SUA levels from the GWAS catalogue (accessed in 07 Feb, 2017). Thirty genetic risk loci in relation to SUA levels ($p<5\times10^{-8}$) were identified from previous GWAS (150, 151). Based on the MR analysis performed by White *et al* (323), one more SNP rs164009 located in the *PRPSAP1* gene ($p=7.06\times10^{-7}$) was included given its functional role in urate metabolism (i.e., it encodes a protein involved in the regulation of purine synthesis). In total, 31 independent SNPs were selected as genetic proxies of SUA level (**Table 4-23**). These SNPs are distributed across 14 chromosomes (Chromosome 1 [n=2], Chromosome 2 [n=3], Chromosome 3 [n=1], Chromosome 4 [n=2], Chromosome 5 [n=1], Chromosome 6 [n=4], Chromosome 7 [n=2], Chromosome 8 [n=2], Chromosome 10 [n=2], Chromosome 11 [n=3], Chromosome 12 [n=2], Chromosome 15 [n=2], Chromosome 16 [n=2], Chromosome 17 [n=3]).

Each SNP effect on SUA level (effect size and standard error [SE]) was taken from the largest meta-analysis of GWAS in European populations performed by the GUGC consortium (151). This meta-analysis included 48 individual GWAS with 110,347 individuals. The overall proportion of variance (adjusted R^2) of SUA level explained by the 31 selected SNPs was estimated to be 7%, of which 3.4% was explained by two SNPs (rs12498742 located at *SLC2A9* and rs2231142 located at *ABCG2*) alone. Genome-wide sex-interaction on the serum urate effect sizes were also examined in this study (151). They performed meta-analyses of GWAS separately for 49,825 men and 60,522 women. Except for *SLC2A9* and *ABCG2*, no additional regions contained SNPs that differed significantly ($p<5\times10^{-8}$) in their effect sizes between men and women. The effect estimates of the 31 selected SNPs on SUA level for the overall and sex-specific subjects are summarised in **Table 4-23**.

Table 4 - 23: The GWAS summary estimates for the effect of 31 SNPs on SUA level.

Genetic instruments					Overall effect			Male-specific effect			Female-specific effect		
SNP	**Chr**	**A1**	**A2**	**Closest/GRAIL gene**	**beta**	**se**	**p_gc**	**beta1**	**se1**	**p1**	**beta2**	**se2**	**p2**
rs10480300	7	T	C	*PRKAG2/PRKAG2*	0.032	0.006	9.37E-07	0.043	0.010	1.70E-05	0.024	0.008	3.20E-03
rs10821905	10	A	G	*A1CF/ASAH2*	0.053	0.007	3.45E-12	0.042	0.011	3.80E-04	0.060	0.009	2.50E-10
rs11264341	1	T	C	*TRIM46/PKLR*	-0.048	0.006	1.04E-14	-0.055	0.009	1.10E-08	-0.044	0.007	9.10E-09
rs1165151	6	T	G	*SLC17A1/SLC17A3*	-0.092	0.005	4.52E-60	-0.096	0.008	1.30E-28	-0.089	0.007	4.20E-37
rs1171614	10	T	C	*SLC16A9/SLC16A9*	-0.074	0.007	6.48E-23	-0.086	0.011	1.90E-13	-0.067	0.009	3.00E-13
rs1178977	7	A	G	*BAZ1B/MLXIPL*	0.050	0.007	6.68E-12	0.055	0.011	8.20E-07	0.046	0.009	2.60E-07
rs12498742	4	A	G	*SLC2A9/SLC2A9*	0.380	0.006	0.00E+00	0.269	0.010	6.40E-153	0.460	0.008	0.00E+00
rs1260326	2	T	C	*GCKR/GCKR*	0.077	0.006	1.31E-40	0.091	0.009	3.00E-25	0.063	0.007	1.90E-04
rs1394125	15	A	G	*UBE2Q2/NRG4*	0.043	0.006	9.78E-11	0.060	0.010	5.50E-06	0.032	0.008	1.00E-04
rs1471633	1	A	C	*PDZK1/PDZK1*	0.061	0.005	1.40E-26	0.069	0.008	3.50E-15	0.054	0.007	1.60E-14
rs164009	17	A	G	*QRICH2/PRPSAP1*	0.029	0.006	7.06E-07	0.024	0.009	6.20E-03	0.032	0.007	8.20E-06
rs17050272	2	A	G	*INHBB/INHBB*	0.037	0.006	9.36E-09	0.049	0.010	6.50E-07	0.030	0.008	1.90E-04
rs17632159	5	C	G	*TMEM171/TMEM171*	-0.038	0.006	2.00E-09	-0.043	0.010	1.30E-05	-0.039	0.008	1.10E-06
rs17786744	8	A	G	*STC1/STC1*	-0.031	0.005	8.82E-08	-0.033	0.009	2.10E-04	-0.029	0.007	2.10E-04
rs2078267	11	T	C	*SLC22A11/SLC22A11*	-0.078	0.006	8.73E-36	-0.085	0.009	2.90E-19	-0.071	0.007	5.70E-20
rs2079742	17	T	C	*BCAS3/C17orf82*	0.051	0.008	6.24E-09	0.054	0.013	5.60E-05	0.048	0.010	1.00E-05
rs2231142	4	T	G	*ABCG2/ABCG2*	0.220	0.009	4.43E-116	0.270	0.014	3.80E-75	0.181	0.011	1.30E-52
rs2307394	2	T	C	*ORC4L/ACVR2A*	-0.035	0.006	7.26E-09	-0.036	0.009	1.20E-04	-0.034	0.007	4.70E-06
rs2941484	8	T	C	*HNF4G/HNF4G*	0.049	0.006	3.91E-17	0.048	0.009	6.20E-08	0.046	0.007	1.30E-10
rs3741414	12	T	C	*INHBC/INHBE*	-0.071	0.007	9.79E-22	-0.091	0.011	7.00E-16	-0.057	0.009	4.30E-10

Genetic instruments					Overall effect			Male-specific effect			Female-specific effect		
SNP	**Chr**	**A1**	**A2**	**Closest/GRAIL gene**	**beta**	**se**	**p_gc**	**beta1**	**se1**	**p1**	**beta2**	**se2**	**p2**
rs478607	11	A	G	*NRXN2/SLC22A12*	-0.048	0.007	5.31E-10	-0.058	0.012	9.60E-07	-0.043	0.009	8.80E-06
rs642803	11	T	C	*OVOL1/LTBP3*	-0.043	0.005	4.51E-14	-0.047	0.008	8.00E-08	-0.042	0.007	2.10E-09
rs653178	12	T	C	*ATXN2/PTPN11*	-0.036	0.005	2.45E-10	-0.044	0.009	7.50E-07	-0.032	0.007	5.50E-06
rs6598541	15	A	G	*IGF1R/IGF1R*	0.044	0.006	5.20E-13	0.039	0.009	2.70E-05	0.050	0.007	1.60E-11
rs675209	6	T	C	*RREB1/RREB1*	0.063	0.006	1.38E-21	0.060	0.010	3.30E-09	0.064	0.008	2.00E-15
rs6770152	3	T	G	*SFMBT1/MUSTN1*	-0.048	0.006	2.66E-16	-0.052	0.009	6.70E-09	-0.047	0.007	6.00E-11
rs7188445	16	A	G	*MAF/MAF*	-0.032	0.006	1.15E-07	-0.025	0.009	7.90E-03	-0.040	0.007	6.40E-08
rs7193778	16	T	C	*NFAT5/NFAT5*	-0.047	0.008	2.36E-08	-0.048	0.012	2.10E-04	-0.045	0.010	1.00E-05
rs7224610	17	A	C	*HLF/HLF*	-0.038	0.006	4.74E-11	-0.043	0.009	9.00E-07	-0.034	0.007	3.00E-06
rs729761	6	T	G	*VEGFA/VEGFA*	-0.046	0.006	3.05E-12	-0.047	0.010	3.20E-06	-0.047	0.008	3.20E-06
rs742132	6	A	G	*LRRC16A/LRRC16A*	0.035	0.006	1.90E-08	0.035	0.006	1.90E-08	0.035	0.006	1.90E-08

4.3.3.2 Use of individual SNPs as genetic instruments

To achieve the primary goal of the PheWAS, which is to identify the cross-phenotype associations (genetic risk locus shared by multiple phenotypes), the simplest and most appropriate approach is to use each single genetic variant of interest as a predictor to observe its associations with the phenome. For this purpose, all 31 SNPs related to SUA level were applied as genetic predictors individually to explore their associations with the phenome. Genotypes of the 31 selected SNPs were extracted from the UK Biobank genetic datasets for the selected study population. Of these, 10 SNPs (rs11264341, rs1260326, rs1394125, rs17050272, rs2078267, rs2231142, rs2307394, rs675209, rs742132, rs653178) were genotyped for all included individuals (n=339,256). Genotypes of the remaining 21 SNPs were extracted from the imputed data. The number of participants with missing genotypes for the 21 SNPs ranges from 51-12,004 (median: 3,206) and the missing rates of genotypic data were less than 5% (range: 0.2%-4.2%; median: 0.09%). The allele frequency and genotype count of the study population (n=339,256) from UK Biobank is presented in **Table 4-24**.

To describe the nature and potential effect of each single variant, the function of putative genes where these variants map to was annotated by using NCBI gene database and *GeneCards* (www.genecards.org) database (*Chapter 1, Section 1.3 "Genetic polymorphisms"* for more information). In summary, of these 31 SNPs, 7 were mapped to genes (*SLC22A11/OAT4, SLC22A12/URAT1, SLC17A1/NPT, PDKZ1, SLC2A9/GLUT9, ABCG2,* and *SLC16A9)* encoding proteins related to urate transport across renal and gut membranes; one SNP (rs164009) is located within a candidate gene (*PRPSAP1*) that encodes a protein involved in the regulation of purine synthesis and thus affects urate generation; the remaining SNPs were largely mapped to genes encoding proteins for transcription or inhibins-activins growth factors with broad downstream responses and highlighting pathways in relation to carbohydrate metabolism, such as regulation of glycolysis, glucose, insulin and pyruvate.

Table 4 - 24: The allele frequency and genotype count of the 31 selected SNPs in the study population (n=339,265) from UK Biobank.

SNP	A1	A2	A1_freq	A2_freq	Genotype 1 (A1A1)	Genotype 2 (A1A2)	Genotype 3 (A2A2)	Missing number	Missing rate
rs10480300	T	C	0.727	0.273	25,212	132,730	178,115	3,206	0.009
rs10821905	A	G	0.824	0.176	10,415	97,284	227,374	4,190	0.012
rs11264341	T	C	0.571	0.429	62,651	166,037	110,575	0	0
rs1165151	T	G	0.549	0.451	68,897	167,982	101,915	469	0.001
rs1171614	T	C	0.769	0.231	17,960	120,040	199,948	1,315	0.004
rs1178977	G	A	0.802	0.198	13,322	107,362	218,528	51	0.000
rs12498742	G	A	0.768	0.232	18,453	120,558	199,922	330	0.001
rs1260326	T	C	0.607	0.393	52,424	161,510	125,329	0	0
rs1394125	A	G	0.638	0.362	44,414	156,803	138,046	0	0
rs1471633	A	C	0.538	0.462	72,514	168,505	97,945	299	0.001
rs164009	G	A	0.613	0.387	50,142	159,289	126,229	3,603	0.011
rs17050272	A	G	0.589	0.411	57,230	164,105	117,928	0	0
rs17632159	C	G	0.697	0.303	31,048	141,910	163,384	2,921	0.009
rs17786744	G	A	0.590	0.410	56,472	162,194	116,523	4,074	0.012
rs2078267	C	T	0.548	0.452	69,205	168,460	101,598	0	0
rs2079742	C	T	0.864	0.136	6,053	78,293	247,447	7,470	0.022
rs2231142	T	G	0.887	0.113	4,470	67,948	266,845	0	0
rs2307394	C	T	0.697	0.303	30,747	144,185	164,331	0	0
rs2941484	T	C	0.553	0.447	65,907	161,615	100,841	10,900	0.032
rs3741414	T	C	0.755	0.245	20,467	124,940	193,647	209	0.001
rs478607	G	A	0.847	0.153	8,044	87,168	243,012	1,039	0.003

SNP	A1	A2	A1_freq	A2_freq	Genotype 1 (A1A1)	Genotype 2 (A1A2)	Genotype 3 (A2A2)	Missing number	Missing rate
rs642803	T	C	0.536	0.464	72,507	166,237	96,539	3,980	0.012
rs653178	C	T	0.517	0.483	79,137	169,471	90,655	0	0
rs6598541	A	G	0.645	0.355	42,133	153,913	139,819	3,398	0.010
rs675209	T	C	0.731	0.269	24,595	133,343	181,325	0	0
rs6770152	G	T	0.576	0.424	60,131	163,329	111,108	4,695	0.014
rs7188445	A	G	0.672	0.328	36,423	148,347	152,201	2,292	0.007
rs7193778	C	T	0.850	0.150	7,559	86,228	243,737	1,739	0.005
rs7224610	C	A	0.604	0.396	51,092	154,994	118,920	14,257	0.042
rs729761	T	G	0.715	0.285	26,764	132,929	167,566	12,004	0.035
rs742132	G	A	0.706	0.294	29,186	140,891	169,186	0	0

4.3.3.3 Calculation of a weighted polygenic risk score as a proxy of SUA level

To explore the potential causal effect of SUA level across phenome-wide health outcomes, a weighted polygenic genetic risk score (GRS) was constructed by incorporating information from the 31 genetic risk loci associated with SUA level. Specifically, the polygenic risk score was created by adding up the number of SUA level increasing alleles carried by all 31 SNPs and then weight it based on their effect estimates on SUA level (regression coefficient beta) derived from the SUA GWAS performed by the GUGC consortium. For instance, if individual i carries g_{ik} copies of the SUA-increasing allele for each variant $k = 1, \ldots, 31$, the weight for variant k is w_k then their weighted polygenic score is $Z_i = \sum_{k=1}^{31} w_k g_{ik}$.

The weighted polygenic risk score was calculated by using plink 2.0 for all individuals (n=339,256) included in the SUA PheWAS analysis. The weighted GRS was normally distributed among the study population with a mean value of 0.436 (SD: 0.309) (**Table 4-25**, **Figure 4-16**).

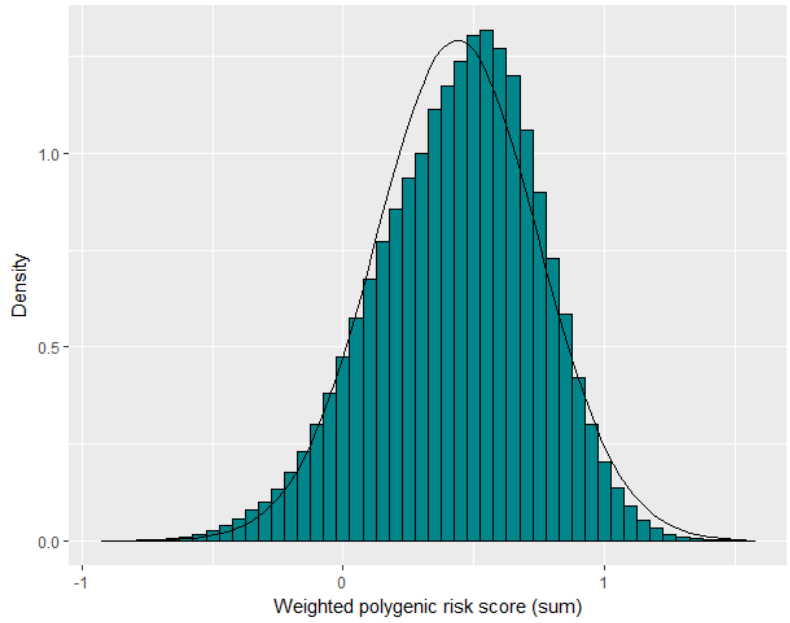


Figure 4 - 16: The distribution of weighted polygenic risk score among study population.

Table 4 - 25: A summary of weighted polygenic risk score among study population.

Weighted GRS	n	Min.	1st Qu.	Median	3rd Qu.	Max.	Mean	SD
	339,256	-0.880	0.231	0.461	0.656	1.545	0.436	0.309

4.3.4 Definition of the phenome framework

The focus on a wider spectrum of phenotypes, termed the phenome, is an important aspect of the PheWAS approach. The phenome has been defined as a systematic and comprehensive set of phenotypes, including both clinical, biochemical and imaging traits, which could be measured as outcome phenotypes (e.g., disease status), or intermediate phenotypes (e.g., clinical variables). Unlike the genome, in which the genetic structure could be measured by reliable biological techniques, generating a framework of phenome highly relies on the clinical measurements and disease diagnoses. The phenome framework defined in currently published PheWASs varies according to studies (the characteristics of these PheWAS are summarised in **Table 4-26**), but the most effective and straightforward way for phenotyping is to use the electronic medical records (EMRs). In this PheWAS analysis for SUA level, I focused on phenotypes in relation to disease outcomes.

4.3.4.1 ICD codes in UK Biobank

We analysed three phenotypic datasets (in-patient hospital records, cancer registry data and death registry data) in the UK Biobank database. The coding for clinical diagnoses in these datasets follows the World Health Organisation's ICD coding systems but uses different ICD versions (ICD-10 or ICD-9) based on the date of record. Within the in-patient hospital data, each episode has a primary ICD-10/9 diagnosis code to describe the event of hospitalisation, and when applicable, one or more secondary diagnosis codes followed to annotate the corresponding hospitalisation event. We included both the primary and secondary codes to define the case and control groups. Since cancer registry and death registry data overlapped with the disease diagnosis in hospital episode records, we pooled them into the hospital episode data to complement the disease diagnoses of participants. In these phenotypic datasets, we identified 2,779,598 unique records of hospital inpatient data corresponding to 395,978 unique individuals (2,714,364 records had an ICD10 diagnosis code, 52,123 had an ICD9 diagnosis code and 13,111 records had no diagnosis code), 233,753 records corresponding to 79,066 unique participants (207,935 records had an ICD10 diagnosis code, and 25,818 had an ICD9 diagnosis code) from the cancer registry data, and 14,417 records from the death registry data. The breadth of the ICD-10/9 coding system ensures that it can describe well the range of human diseases but the individual ICD codes cannot be directly used to define an independent phenotype, as they are not designed to represent distinct phenotype groups. To aggregate the ICD codes, the PheCODE schema has been developed and successfully adopted in a number of PheWAS to combine one or more individual ICD codes into distinct phenotype groups (see the *Chapter 4, Section 4.3.4.2 "PheCODE*

schema" for more information) (433-435).

Table 4 - 26: The characteristics of currently published PheWAS studies.

Study	Population	Sample size	Selected Predictors	Phenotype data source	No. of Phenotypes	Significant phenotypic associations
Genetic variants						
Denny, 2010 (436)	BioVU	6005	5 SNPs (with known associations)	EMR (ICD-9-CM)	733	23 phenotypes (4 replicated, 19 novel)
Denny, 2011 (437)	eMERGE	13,617	1 SNP (near *FOXE1*)	EMR(ICD-9-CM)	866	Hypothyroidism, other thyroid diseases
Pathak, 2012 (438)	Mayo Clinical Biobank	6307	4 SNPs (associated with T2DM and related traits)	EMR(ICD-9-CM)	285	Diabetes, disorders of lipid metabolism
Denny, 2013 (439)	eMERGE	21,241	3144 SNPs (all SNPs in GWAS catalog)	EMR(ICD-9-CM)	1385	263 phenotypes (210 replicated with $p<0.05$, 63 novel FDR<0.1)
Hebbring, 2013 (440)	Marshfield Clinical patients	4235	1 SNP (tagging HLA-DRB1*1501)	EMR(ICD-9-CM)	4841	MS, alcohol–induced cirrhosis, erythematous conditions, benign neoplasms of respiratory organs
Pendergrass, 2013 (441)	PAGE network	70,061	83 SNPs (SNPs overlapped across PAGE study sites)	Cohort studies	4706	111 phenotypes (26 replicated, 33 novel)
Ritchie, 2013 (434)	eMERGE	13,859	23 SNPs (associated with ECG QRS duration)	EMR(ICD-9-CM)	778	Cardiac arrhythmias; atrial fibrillation
Cronin, 2014 (442)	Emerge & BioVU	24,198	54 SNPs (all in *FTO* gene)	EMR(ICD-9-CM)	1645	Obesity, T2DM, fibrocystic breast disease
Hall, 2014 (443)	Genetic NHANES	14,042	80 SNPs (GWAS-identified SNPs in NHANES datasets)	Cohort studies	1008	69 phenotypes (21 novel)
Mitchell, 2014 (444)	BioVU	11,519	130 SNPs (cardiovascular-related mtSNPs)	EMR derived	8	T2DM, total cholesterol level
Namjou, 2014 (445)	Paediatric patients	4268	2476 SNPs (overlapped SNPs between the dataset and GWAS catalog)	EMR(ICD-9-CM)	539	JRA, thyroiditis, T1DM
Shameer, 2014 (446)	eMERGE	13,582	81 SNPs (associated with platelet parameters)	EMR(ICD-9-CM)	1368	Myocardial infarction, autoimmune, hematologic disorders
Carroll, 2014 (447)	BioVU	6005	1 SNP (near *HLA-DRB*, associated with multiple sclerosis)	EMR(ICD-9-CM)	1127	Multiple sclerosis
Ye, 2015 (448)	Marshfield Clinical patients	14,875	105 SNPs (presumed functional stop-gain and stop-loss variants)	EMR(ICD-9-CM)	4841	Age-related macular degeneration
Diogo, 2015 (449)	BioVU & other	29,377	3 SNPs (in *TYK2* gene)	EMR(ICD-9)	502	Rheumatoid Arthritis (RA)

Study	Population	Sample size	Selected Predictors	Phenotype data source	No. of Phenotypes	Significant phenotypic associations
Moore, 2015 (450)	AIDS Clinical Trial Group	2547	10,584 SNPs (SNPs passed the filter threshold across datasets)	Clinical trial (Laboratory Phenotypes)	27	Higher total bilirubin Lower absolute neutrophil counts LDL-C
Hebbring, 2015 (451)	Marshfield Clinical patients	4235	5 SNPs (reported in GWAS)	EMR (ICD code & text-mining)	23,384 & 4841	Nonexudative senile macular degeneration, spondylitis, MS, atrial fibrillation, triglyceride level
Non-genetic predictors						
Warner, 2012 (452)	MIMIC II	36,095	Peak WBC count	EMR(ICD-9-CM)	5756	26 phenotypes (e.g. unspecified septicaemia)
Boland, 2013 (453)	New Yorkers	2475	Presence of periodontitis	EMR(ICD-9-CM)	993	9 phenotypes (e.g. T1DM, T2DM)
Liao, 2013 (454)	RA cases vs. controls	2526	Autoantibody	EMR(ICD-9-CM)	512	Hypothyroidism, sicca, chronic non-alcoholic liver disease
Neuraz, 2013 (455)	Patients treated by thiopurine drugs	442	TPMT activity	EMR(ICD-10)	445	Diabetes, nutritional anaemia
Warner, 2013a (456)	MIMIC II	24,580	Length of hospitalisation	EMR(ICD-9-CM)	5657	191 (e.g. hospital-acquired complications)
Warner, 2013b (457)	MIMIC II & SHRINE	3392	Myeloma	EMR(ICD-9-CM)	431	24 (e.g. Pathologic fraction)
Roesch, 2015 (458)	Geisinger Clinic's bariatric surgery program	128	*FGF19* and *FGF21*	EMR	205	Higher glucose

4.3.4.2 PheCODE schema

The PheCODE schema was developed to combine one or more related ICD codes into distinct diseases or traits (459). In principle, ICD codes representing diseases with common aetiology were combined into one phenotype group defined by the phecode. On the other hand, for some diseases with distinctly different aetiologies, like type 1 and type 2 diabetes, their individual codes are retained and allocated into different phenotype groups. The PheCODE system has been continuously refined by clinical experts helping with revisions of different domains, such as cardiology and oncology. The latest version of the PheCODE system includes 1,866 hierarchical phenotype codes. The PheCODE system also provides a scheme to automatically exclude patients that have similar or potentially overlapping diseases (e.g. excluding type 1 diabetes and secondary diabetes mellitus for an analysis of type 2 diabetes) from the corresponding control group. Although the PheCODE system is effective at replicating genotype-phenotype associations, the current version of the PheCODE system was designed for the International Classification of Diseases, version 9, Clinical Modification (ICD-9-CM), and the phecode algorithm was not directly applicable to the ICD-10 coding system in the UK Biobank.

To develop an aggregation method for the PheWAS analysis applicable to the ICD-10 coding system in UK Biobank, we collaborated with the Electronic Medical Records and Genomics (eMERGE) group of Vanderbilt University Medical Center and mapped ICD-10 codes to phecodes in both direct and indirect ways (**Figure 4-17**). An ICD-10 code could be mapped to a phecode directly if their descriptions matched each other regardless of capitalisation. An ICD-10 code can also be mapped to a phecode indirectly through an ICD-9-CM code. The unified medical language system (UMLS) was used to map the ICD-10 code to ICD-9-CM (or map the ICD-10 code to systematised nomenclature of medicine clinical terms [SNOMED CT] code first and then to ICD-9-CM) and then the previous mapping of ICD-9-CM to phecode was used to finally link the ICD-10 to the phecode. When an ICD-10 code could be mapped to both a child phecode and its parent phecode, only the mapping to the child phecode was retained and the child phecode was then mapped to its parent phecode; or when an ICD-10 code can be mapped to multiple distinct phecodes, all the mappings are kept. The mapping process was developed by the Vanderbilt group and I matched the ICD codes in the UK Biobank to the PheCODE schema for analysis.

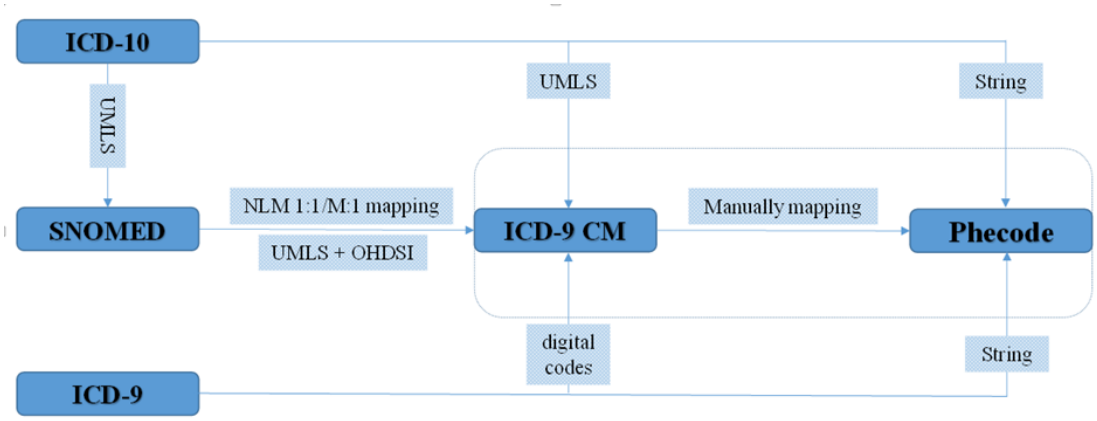


Figure 4 - 17: Mapping ICD-10 codes to phecodes.

Abbreviations: SNOMED CT = Systematised Nomenclature of Medicine, Clinical Terms; UMLS = Unified Medical Language System; GEM = General Equivalence Mappings; OHDSI = Observational Health Data Sciences and Informatics.

Among the 7,990 ICD-10 codes used in UKBB, 6,654 (83.2%) codes were successfully mapped to phecodes. An additional 3,500 ICD codes within UKBB did not exist in the current ICD-10 versions, therefore, were not mapped to phecodes. When examining the mapping procedure, the majority of the unmapped ICD-10 codes were composed of encounter or procedural codes (i.e. codes beginning with Z) or supplementary codes (i.e. codes beginning with Y), which were not a description of a specific phenotype or disease. In addition, some ICD-10 code areas used in the UKBB representing the personal and family history were also unable to be mapped due to the missing corresponding elements in the PheCODE system. The top 10 unmapped ICD-10 codes are listed in **Table 4-27**. In total, 8,947 ICD-10 codes were mapped to at least one phecode, with 256 (2.9%) codes mapped to more than one phecode. After mapping diagnostic ICD-10/9 codes to phecodes, the phenotypic data of UK Biobank consisted of 1807 distinct phecodes.

Table 4 - 27: The top ten unmapped ICD-10 codes in the UKBB.

	UKBB ICD-10	ICD-10 descriptions
In-hospital and cancer registry data	Z53.8	Procedure not carried out for other reasons
	Z92.2	Personal history of long-term (current) use of other medicaments
	Z30.2	Sterilisation
	Z82.4	Family history of ischaemic heart disease and other diseases of the circulatory system
	Z37.0	Single live birth
	Z03.8	Observation for other suspected diseases and conditions
	Z12.1	Special screening examination for neoplasm of intestinal tract
	Z92.1	Personal history of long-term (current) use of anticoagulants
	Z96.6	Presence of orthopaedic joint implants
	Z09.0	Follow-up examination after surgery for other conditions

4.3.4.3 Tree-structured phenotypic model

A novel Bayesian analysis framework, termed a tree-structured phenotypic model (TreeWAS), was developed to interrogate increasingly specific sub-phenotypes encoded by ICD-10 codes while retaining the statistical power to detect genetic associations (460). This application models the genetic coefficients across all phenotypes as a set of random variables. Given that ICD-10 codes are organised in a hierarchical tree-like structure, to model the correlation of this structure, a Markov process is applied to allow the genetic coefficients to evolve down the tree trunk and branches. The tree structure is determined from a known classification hierarchy based on the ICD coding system where each node of the tree is a clinical term in the classification and observations can be made at both terminal and internal nodes. The prior probability determines the expected degree of correlation between genetic coefficients across phenotypes. The coefficient at a parent node can either be inherited by a child node (the probability is denoted as $e^{-\theta}$) or can transition to a new uncorrelated value (the probability is denoted as $1- e^{-\theta}$). This new value will be zero with a probability of $1 - \pi_1$, or non-zero with a probability π_1. Thus, the $e^{-\theta}$ and π_1 parameters define the transition probabilities that control the Markov process. Given the structure of the model and the Markov process assumption, the likelihood over the genetic coefficients could be calculated across all clinical phenotypes using a dynamic programming model and the forward and backward algorithms (see reference (460) for more details). An overall Bayes Factor (BF) is estimated if the genetic coefficients are non-zero for at least one of the nodes in the tree. The marginal posterior probability (PP) at each node in the tree where the genetic

coefficient is non-zero, and the magnitude of the corresponding effect were determined by using the maximum a posteriori (MAP) estimator (see reference (460) for more details).

5 MR-PheWAS: INTERIM RELEASE OF UK BIOBANK DATA

5.1 Summary

This chapter presents a MR-PheWAS (phenome-wide association study incorporated with Mendelian randomisation [MR]) analysis that was performed to investigate the associations between the 31 SUA genetic risk variants and a very wide range of disease outcomes (n=568) by using the interim release data of UK Biobank (n=120,091). The SUA genetic risk loci were employed as instruments individually. The framework of phenome was defined by the PheCODE schema using the ICD diagnosis codes documented in the health records of UK Biobank. Phenome-wide association tests were performed first to identify associations across the SUA genetic risk loci and the Phenome; MR and HEIDI (heterogeneity in dependent instruments) analyses were then applied to investigate whether the observed PheWAS associations were due to causality, pleiotropy or genetic linkage.

The MR-PheWAS (using the interim release cohort) identified 25 disease groups/ outcomes to be associated with SUA genetic risk loci after multiple testing correction ($p<8.57\times10^{-5}$). The MR IVW (inverse variance weighted) analysis implicated a causal role of SUA level in three disease groups: inflammatory polyarthropathies (OR=1.22, 95% CI: 1.11 to 1.34), hypertensive disease (OR=1.08, 95% CI: 1.03 to 1.14) and disorders of metabolism (OR=1.07, 95% CI: 1.01 to 1.14); and four disease outcomes: gout (OR=4.88, 95% CI: 3.91 to 6.09), essential hypertension (OR=1.08, 95% CI: 1.03 to 1.14), myocardial infarction (OR=1.16, 95% CI: 1.03 to 1.30) and coeliac disease (OR=1.41, 95% CI: 1.05 to 1.89). After balancing pleiotropic effects in MR Egger analysis, only gout and its encompassing disease group of inflammatory polyarthropathies were considered to be causally associated with SUA level. The analysis also highlighted a locus (*ATXN2/S2HB3*) that may influence SUA level and multiple cardiovascular and autoimmune diseases via pleiotropy.

This chapter has been published in the *Annals of the Rheumatic diseases* cited as "*Li, X., Meng, X., Spiliopoulou, A., Timofeeva, M., Wei, W.Q., Gifford, A., Shen, X., He, Y., Varley, T., McKeigue, P., Tzoulaki, I., McKeigue, P., Joshi, P., Denny, J.C., Campbell, H. and Theodoratou, E. 2018. MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank. Annals of the rheumatic diseases, pp.annrheumdis-201. doi: 10.1136/annrheumdis-2017-212534" (461).*

I was fully involved in all aspects of the research work presented in this chapter and the publication. In the study design, I read a vast amount of literature and found different

methodologies to decide the most appropriate one for this study. During the implementation of data analysis, I sought guidance from my supervisors to ensure each step of the analysis was robust and reasonable. I accomplished the data analysis with detailed reports to present the research findings. For publishing this work, I wrote the manuscript and revised the paper according to the comments given by the peer reviewers and the journal editors. Specific to the contribution of co-authors, *Theodoratou, E.* and *Campbell, H.* conceived the study. *Meng, X., Wei, Q., Gifford, A., Tzoulaki, I., Denny, J.C., and Varley, T.,* contributed to create the mapping of ICD-10/9 codes to phecode. All authors critically reviewed the manuscript and contributed important intellectual content.

MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank

AUTHORS Xue Li[1*], Xiangrui Meng[1*], Athina Spiliopoulou[1], Maria Timofeeva[2], Wei-Qi Wei[3], Aliya Gifford[3], Xia Shen[1,4], Yazhou He[1,5], Tim Varley[6], Paul Mckeigue[1], Ioanna Tzoulaki[7,8,9], Alan F. Wright[10], Peter Joshi[1], Joshua C. Denny[3], Harry Campbell[1], Evropi Theodoratou[1,11]

1 Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

2 Colon Cancer Genetics Group, Medical Research Council Human Genetics Unit, Medical Research Council, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

3 Department of Biomedical Informatics, Vanderbilt University Medical Centre, Nashville, TN, USA

4 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

5 West China School of Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan Province, P.R. China

6 Public Health and Intelligence, NHS National Services Scotland, Edinburgh, UK

7 Department Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

8 MRC-PHE Centre for Environment, School of Public Health, Imperial College London, London, UK

9 Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece

10 Medical Research Council, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

11 Edinburgh Cancer Research Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

[*] The authors contribute equally to this work.

ABSTRACT

Objectives: We aimed to investigate the role of serum uric acid (SUA) level in a broad spectrum of disease outcomes using data for 120,091 individuals from UK Biobank.

Methods: We performed a phenome-wide association study (PheWAS) to identify disease outcomes associated with SUA genetic risk loci. We then implemented conventional Mendelian randomisation (MR) analysis to investigate the causal relevance between SUA level and disease outcomes identified from PheWAS. We next applied MR Egger analysis to detect and account for potential pleiotropy, which conventional MR analysis might mistake for causality, and used the HEIDI (heterogeneity in dependent instruments) test to remove cross-phenotype associations that were likely due to genetic linkage.

Results: Our PheWAS identified 25 disease groups/outcomes associated with SUA genetic risk loci after multiple testing correction ($p<8.57\times10^{-5}$). Our conventional MR analysis implicated a causal role of SUA level in 3 disease groups: inflammatory polyarthropathies (OR=1.22, 95%CI: 1.11 to 1.34), hypertensive disease (OR=1.08, 95%CI: 1.03 to 1.14), disorders of metabolism (OR=1.07, 95% CI: 1.01 to 1.14), and 4 disease outcomes: gout (OR=4.88, 95%CI: 3.91 to 6.09), essential hypertension (OR=1.08, 95%CI: 1.03 to 1.14), myocardial infarction (OR=1.16, 95%CI: 1.03 to 1.30), coeliac disease (OR=1.41, 95%CI: 1.05 to 1.89). After balancing pleiotropic effects in MR Egger analysis, only gout and its encompassing disease group of inflammatory polyarthropathies were considered to be causally associated with SUA level. Our analysis highlighted a locus (*ATXN2/S2HB3*) that may influence SUA level and multiple cardiovascular and autoimmune diseases via pleiotropy.

Conclusions: Elevated SUA level is convincing to cause gout and inflammatory polyarthropathies, and might act as a marker for the wider range of diseases with which it associates. Our findings support further investigation on the clinical relevance of SUA level with cardiovascular, metabolic, autoimmune, and respiratory diseases.

5.2 Introduction

Uric acid (UA) is the end product of the exogenous and endogenous purine metabolisms, catalysed by the action of xanthine oxidase (8). Due to the evolved loss of uricase enzyme, humans are unable to convert UA into highly soluble compounds, leaving urate circulating in blood and resulting in a high basal level of serum uric acid (SUA).(34) The prevalence of hyperuricaemia (elevated SUA level >416 μmol/L) is in the range of 5-25% across different countries (91, 462, 463). A progressively rising trend of hyperuricaemia prevalence has been observed worldwide (463). Concerningly, hyperuricaemia is thought to inflict multiple clinical consequences, which is believed to be causally related to gout and suggestively associated with a number of prevalent health conditions, such as cardiovascular and metabolic diseases (236, 464, 465).

Our recently published umbrella review presented a comprehensive overview of the breadth of disease outcomes related to SUA level by incorporating evidence from multiple sources (466). A large number of disease outcomes were reported to be associated with SUA level in observational studies, covering a wide range of diseases, including cardiovascular disease, metabolic syndrome, diabetes, cancer, and neurological disorders. However, evidence as to whether these associations are actually causal is not yet well developed, given that observational associations are susceptible to a variety of biases, confounding and/or reverse causality. Although results from randomized controlled trials (RCTs) have provided some evidence about the beneficial effects of SUA-lowering therapy on some intermediate traits or biomarkers (e.g. blood pressure, endothelial function, serum creatine), there remains a lack of RCTs focusing on the more important clinical disease endpoints (50, 316, 320). A number of Mendelian randomisation (MR) studies, using the genetic variants influencing SUA level as instruments, provide alternative evidence to distinguish causal from non-causal associations. However, these MR studies examined a limited set of disease outcomes and were not able to detect moderate effect size due to limited power (323, 326-328, 332, 337, 349). Increasing sample size and the range of outcomes in an enlarged MR study thus offers the prospect of deeper and wider insight into the causal role of SUA.

MR analysis is typically hypothesis-driven based on prior knowledge to specify the outcome to be examined in relation to the exposure of interest. Traditionally, only one (or a limited number) association between the exposure and one (or a few) pre-defined outcome(s) is tested in a MR study. Recently, Phenome-wide Mendelian randomisation (MR-PheWAS) analysis is proposed by integrating the Phenome-wide association study (PheWAS) and MR

method to build a hypothesis-searching approach, which aims to explore potential causal relationships between an exposure (using genetic instruments as proxies) and a range of phenome-wide disease outcomes in a high-throughput manner (467). This approach is effective in evaluating or replicating the associations reported in observational studies, as well as discovering new relationships, and generating new hypotheses on the genetic architecture shared by the related phenotypes. With its wealth of genotypic and phenotypic data collected in very large numbers, the UK Biobank study provides an excellent opportunity to explore the causal role of SUA level across a broad spectrum of disease outcomes. In this study, we performed a MR-PheWAS in UK Biobank database to discover disease outcomes related to genetic variations of SUA level and to investigate if any association is causal.

5.3 Methods

5.3.1 UK Biobank data

The UK Biobank is a large-scale population-based prospective cohort, which enrolled over 500,000 participants aged at 40-69 years. The recruited participants provided a wide range of self-reported baseline information. Blood samples were collected for biochemical tests and genotyping. Their national health records has been linked with the baseline and genotypic data for longitudinal follow-up. Genotypic and phenotypic data used in this study were obtained from UK Biobank under an approved data request application (application ID: 10775).

5.3.2 Genotyping and quality control

Genotyping, quality control and genetic imputation were performed by the UK Biobank team prior to the interim release of genotypic data for 150,000 participants. The procedure of genotyping and quality control is presented in detail at https://biobank.ctsu.ox.ac.uk/crystal/docs/genotyping_qc.pdf. We used the field variables made available by the UK Biobank for quality control to exclude the samples that had high missing or heterozygosity, outlying short runs of homozygosity, and sex mismatch (see online **Supplementary Table 5-1**). We constrained our analyses to participants who were self-reported British and confirmed to be European heritage based on the genetic principle component analysis performed by the UK Biobank. The quality control process generated a genotypic dataset output with 120,091 individuals included in the current analysis.

5.3.3 Phenotyping and mapping ICD-10/9 to phecode

We focused on phenotypes in relation to diagnostic disease outcomes. We analysed two phenotypic datasets (in-patient hospital episode records and cancer registry data) in the UK Biobank using the PheCODE schema (see online supplementary text for phenotyping and mapping process).(433) The coding for clinical diagnoses in these datasets followed the World Health Organisation's International Classification of Diseases (ICD) coding systems, but used different ICD versions (ICD-10 or ICD-9) according to the date of record. We included both ICD-10 and ICD-9 codes to define the case and control groups. Since cancer registry data overlapped with the cancer diagnosis in in-patient hospital records, we pooled the cancer registry data into the hospital episode data as a complement to the cancer diagnosis.

5.3.4 Statistical analysis

The statistical analysis included three main steps: first, we performed a PheWAS to identify disease outcomes that were associated with genetic risk loci of SUA level; second, we performed MR analysis by using both the inverse-variance weighted (IVW) method and MR Egger approach to explore causal relationship for identified PheWAS associations (468, 469); third, we applied HEIDI (heterogeneity in dependent instruments) test to exclude the cross-phenotype associations caused by genetic linkage (470).

5.3.4.1 Genetic instruments

We selected 31 SUA-associated SNPs as genetic instruments (see online **Supplementary Table 5-2**), which were previously reported to be independently associated with SUA level in genome-wide association studies (GWAS) (150, 151). We obtained the SNP effect on SUA level from the largest GWAS performed in European population (151). The overall proportion of variance (adjusted R^2) of SUA level explained by the selected genetic instruments was estimated to be close to 7.0% (151).

5.3.4.2 Phenome-wide association analysis

In phenome-wide analysis, we used 31 SUA-associated SNPs as genetic instruments individually to scan across a wide range of disease outcomes defined by the PheCODE system (433). With the PheWAS algorithm (471), a series of PheWAS association tests were performed: (i) the case group was generated by including patients with the tested phecode; (ii) participants were assigned to the control group based on the absence of both the tested

phecode and related phecodes (patients who had the parent, child or sibling phecodes of the tested phecode were excluded from the control group) (471); (iii) to ensure statistical power, analysis was only performed for phecode with no less than 200 cases. This minimum number of cases was suggested based on a simulation of power estimates for PheWAS analysis (472). We used logistic regression to test the associations between 31 individual genetic instruments (assuming an additive genetic model) and each phecode (number of cases ≥ 200) after adjusting for multiple covariates, including sex, BMI, age, assessment center and the principle components. Considering many phecodes were not independent, we used the false discovery rate (FDR) method developed by Benjamini *et al* to account for multiple testing (473).

5.3.4.3 MR IVW, MR Egger and HEIDI test

We then explored the identified PheWAS associations in three possible scenarios (see online **Supplementary Figure 5-1**): (i) causality: the observed association was causal (through the SUA pathway); (ii) pleiotropy: the observed association was due to pleiotropic effect of one causal variant, (i.e. linked to SUA level and the particular disease outcome through pleiotropy); (iii) genetic linkage: the observed association was caused by the LD between two distinct causal variants, with one affecting SUA level and the other affecting the disease outcome.

MR IVW To explore if there was any causal effect on identified disease outcomes, we performed the conventional MR analysis by pooling the individual effect of each SNP using the IVW method to estimate the overall causal effect (see online supplementary text) (474).

MR Egger We then performed MR Egger to attempt to correct for any potential pleiotropic effect in the causal estimates. This approach is applied to balance the pleiotropic effects derived from multiple genetic instruments (see online supplementary text) (469).

HEIDI test We calculated HEIDI statistics for the SUA genetic loci that were associated with more than one disease outcome. This test was to examine if the cross-phenotype association was due to genetic linkage (see online supplementary text) (470).

5.3.5 Sex stratification analysis

To account for any sex difference, we performed PheWAS and MR analyses in men and women separately. The sex-specific effects of SNPs on SUA level (see online **Supplementary Table 5-2**) were taken from the summary-level GWAS data provided by Köttgen *et al* (151).

5.4 Results

A total of 120,091 UK Biobank participants were included in the analysis, consisting of 56,845 men and 63,246 women with a mean age of 64.86 years in 2016 (standard deviation [SD] of 7.95) (see online **Supplementary Table 5-3**). Within phenotypic datasets, we identified 684,324 hospital episodes and 23,174 cancer registration records, which included 7,990 unique ICD-10 codes and 1,998 unique ICD-9 codes. After mapping diagnostic ICD-10/9 codes to phecodes, the phenotypic data consisted of 1807 distinct phecodes. After filtering out disease outcomes with low prevalence (number of cases <200), 568 phecodes (median number of cases = 694 [range: 200-39,142]) were included in PheWAS analysis. These 568 phecodes were classified into 17 broadly related disease categories (**Table 5-1**). We noted that the distribution of phenotypes examined was skewed across the different disease categories (see online **Supplementary Figure 5-2**), in which a large number of disease phenotypes was included in digestive, circulatory, endocrine and metabolic systems, but some disease categories, for example congenital anomalies, were not well represented in the study population.

Table 5 - 1: The number of phenotypes and the number of cases in each disease category.

Disease categories	Number of phenotypes	Number of cases			
		Min.	Median	Mean	Max.
Circulatory system	61	221	665	2,937	39,142
Congenital anomalies	6	206	265	302	522
Dermatological diseases	24	201	706	2,736	32,738
Diseases in sense organs	34	201	425	1,216	11,306
Digestive diseases	73	201	949	2,176	23,129
Neoplasms	59	203	763	1,916	30,101
Infectious diseases	16	205	787	975	3,192
Endocrine and metabolic diseases	25	229	492	2,304	13,592
Hematopoietic diseases	10	205	1,187	1,600	3,669
Neurological diseases	21	229	452	1,282	11,828
Respiratory diseases	38	219	712	1,713	19,238
Mental disorders	18	205	673	1,926	8,942
Genitourinary diseases	77	200	666	1,606	29,859
Pregnancy complications	11	227	360	707	2,531
Musculoskeletal diseases	44	263	1,076	2,482	21,822
Clinical symptoms	14	267	1,237	2,570	12,287
Injuries and poisonings	37	211	589	911	4,842

5.4.1 Phenome-wide association analysis

The PheWAS analysis performed 17,608 case-control tests, leading to an adjusted significance threshold of $p<8.57\times10^{-5}$ corresponding to a FDR of $q<0.05$ to account for the multiple testing. A total of 27 pairs of genotype-phenotype associations passed the significance threshold of FDR correction ($p<8.57\times10^{-5}$) in the overall PheWAS analysis with adjustment for covariates (**Table 5-2**). Results of PheWAS without adjustment for BMI are shown in online **Supplementary Table 5-4**. The sex-stratified PheWAS analysis identified 10 pairs of genotype-phenotype association in men and 10 pairs of genotype-phenotype association in women (see online **Supplementary Table 5-5**). When compared to the overall PheWAS analysis, 5 new pairs of association were identified from the sex-stratified PheWAS analysis (see online **Supplementary Table 5-5**).

These identified genotype-phenotype associations were distributed across 15 SUA genetic loci, of which 5 loci were associated with more than one disease outcome: rs653178 in the *ATXN2/SH2B3* locus (number of disease outcomes: $n_{outcomes} = 10$), rs1165151 in the *SLC17A3* locus ($n_{outcomes}$ =3), rs1260326 in the *GCKR* locus ($n_{outcomes}$ =3), rs2231142 in the *ABCG2* locus ($n_{outcomes}$ =4) and rs2079742 in the *BCAS3* locus ($n_{outcomes}$ =2). Of note, six disease outcomes shared genetic associations with SUA level at more than one locus: gout (number of loci: n_{loci} =3), inflammatory polyarthropathies (n_{loci} =2), disorders of iron metabolism (n_{loci} =2), coeliac disease (n_{loci} =2), hypertensive disease (n_{loci} =2) and essential hypertension (n_{loci} =2).

In summary, the PheWAS analyses identified 25 unique disease groups/outcomes (corresponding to 25 unique phecodes) that shared genetic risk loci with SUA level, which included 9 disease groups (inflammatory polyarthropathies, hypertensive disease, circulatory disease, disorders of metabolism, disorders of thyroid, other diseases of respiratory system, disorder of skin and subcutaneous tissue, benign neoplasm of digestive system, and complications of labor and delivery) and 16 specific disease outcomes (gout, essential hypertension, angina pectoris, myocardial infraction, coronary atherosclerosis, ischaemic heart disease, atrial fibrillation and flutter, varicose veins of lower extremity, hypercholesterolaemia, disorders of iron metabolism, coeliac disease, hypothyroidism, gastritis and duodenitis, poisoning by antibiotics, cataract, and nasal polyps). The mappings of ICD codes to these 25 phecodes and their hierarchical relationships are shown in online **Supplementary Table 5-6**.

5.4.2 MR IVW, MR Egger and HEIDI test

We then performed MR analysis using the IVW method to explore if there was any causal link between SUA level and the 25 disease groups/outcomes identified from PheWAS analysis. The MR IVW analysis suggested a potential causal link for 7 out of 25 disease groups/outcomes. The corresponding effect estimate on each disease outcome is presented in **Table 5-3**. It was indicated that genetically determined higher SUA level was potentially causally linked with an increased risk of 3 disease groups, including inflammatory polyarthropathies (OR=1.22, 95% CI: 1.11-1.34, p=1.10×10^{-4}), hypertensive disease (OR=1.08, 95% CI: 1.03-1.14, p=0.004), disorders of metabolism (OR=1.07, 95% CI:1.01-1.14, p=0.03), and of 4 specific disease outcomes including gout (OR=4.88, 95% CI: 3.91-6.09, p=3.55e-15), essential hypertension (OR=1.08, 95 %CI: 1.03-1.14, p=0.005), myocardial infarction (OR=1.16, 95% CI: 1.03-1.30, p=0.015), coeliac disease (OR=1.41, 95% CI:1.05-1.89, p=0.02).

To explore and correct for any possible pleiotropic effect of multiple instruments, we then conducted the MR Egger analysis (**Table 5-3**). After balancing out the potential pleiotropic effects, the putative causal link of SUA level with gout (OR=4.58, 95%CI: 2.72 to 7.72, P_{effect} = 1.76×10^{-6}) and its umbrella disease group, inflammatory polyarthropathies (OR=1.15, 95%CI: 1.01 to 1.31, P_{effect}=0.03) remained statistically significant and there was no indication of unbalanced pleiotropy ($P_{pleiotropy}$ =0.73 and $P_{pleiotropy}$ =0.23, respectively). The putative causal effect of SUA level on the other 5 disease groups/outcomes was not statistically significant in the MR Egger model. The causal effects of each individual SNPs on these 7 disease groups/outcomes are shown in online **Supplementary Figures 5-3**, **5-4**, **5-5**, **5-6**, **5-7**, **5-8**, and **5-9**. Unbalanced pleiotropy was observed for essential hypertension ($P_{pleiotropy}$=0.001) and its umbrella disease group, hypertensive disease ($P_{pleiotropy}$=0.001). For myocardial infarction, coeliac disease and disorders of metabolism, the putative causal effect was not statistically significant in the MR Egger model (P_{effect}=0.75, P_{effect} =0.41 and P_{effect} =0.80, respectively), although there was no evidence of unbalanced pleiotropy ($P_{pleiotropy}$ =0.13, $P_{pleiotropy}$ =0.75 and $P_{pleiotropy}$=0.18, respectively). The results of the sex-stratified MR IVW are presented in online **Supplementary Table 5-7.**

Finally, to distinguish the genotype-phenotype association of pleiotropy from LD, the HEIDI test was performed for the 5 genetic loci (rs653178 at *ATXN2/SH2B3*, rs1165151 at *SLC17A3*, rs1260326 at *GCKR*, rs2231142 at *ABCG2* and rs2079742 at *BCAS3*) that were

associated with multiple disease outcomes in the PheWAS analysis (see online **Supplementary Figures 5-10**, **5-11**, **5-12**, **5-13**, and **5-14**). Based on the HEIDI test, we identified 14 disease outcomes that were associated with the SUA genetic risk loci due to pleiotropy (with P_{HEIDI} >0.05). The strongest pleiotropic locus was the *ATXN2/SH2B3*, where three SNPs (rs653178, rs4766578, and rs3184504) in near-complete LD (r^2=0.99) were tagged as the lead SNPs associated with 10 disease groups/outcomes as a cluster of cardiovascular diseases and autoimmune disorders (see online **Supplementary Figure 5-10**). Other potential pleiotropic effects included the associations of the *BCAS3* locus (rs2079742) with essential hypertension (P_{HEIDI} =0.10) and hypertensive disease (P_{HEIDI} =0.09) (see online **Supplementary Figure 5-11**), the associations of the *ABCG2* locus (rs2231142) with varicose veins of lower extremity (P_{HEIDI} =0.32) (see online **Supplementary Figure 5-12**), and the association of the *SLC17A3* locus (rs1165151) with poisoning by antibiotics (P_{HEIDI} =0.26) (see online **Supplementary Figure 5-13**).

Our analysis rejected the null hypothesis of a pleiotropic model for the shared genetic association between SUA level and disorders of iron metabolism at the *SLC17A3* locus (rs1165151) (P_{HEIDI} =5.54×10^{-28}); we identified a different causal variant (rs17342717 in *SLC17A1*) that was in LD with the SNP rs1165151 (r^2=0.24) and strongly associated with the disorders of iron metabolism (P=1.69×10^{-129}) (see online **Supplementary Figure 5-13**). Similarly, for the associations between the *SLC17A3* locus (rs1165151) and coeliac disease (P_{HEIDI} =6.51×10^{-16}) (see online **Supplementary Figure 5-13**), the *GCKR* locus (rs1260326) and hypercholesterolaemia (P_{HEIDI} =3.27×10^{-11}) (see online **Supplementary Figure 5-14**), the pattern of shared regional genetic association was more consistent with a genetic linkage model, and the SNP with the smallest p-value was tagged as an index of the distinct causal variant affecting the examined disease outcome.

Table 5 - 2: Genotype-phenotype associations identified from PheWAS after correcting multiple testing by FDR ($p<8.57\times10^{-5}$).

Phecode	Description	SNP_risk allele†	allele_freq	n_total	n_cases	OR (95%CI)	P*
274.1	Gout	rs2231142_T	0.11	119,555	1,003	1.89 (1.69, 2.12)	5.41e-28
275.1	Disorders of iron metabolism	rs1165151_G	0.45	119,063	205	3.56 (2.78, 4.56)	1.41e-23
244.4	Hypothyroidism	rs653178_C	0.48	118,821	4,146	1.21 (1.16, 1.27)	3.90e-17
246	Disorders of thyroid	rs653178_C	0.48	119,601	4,926	1.18 (1.14, 1.23)	8.82e-16
274.1	Gout	rs12498742_A	0.23	118,960	1,002	1.54 (1.37, 1.74)	7.94e-13
275.1	Disorders of iron metabolism	rs742132_A	0.29	119,271	205	2.80 (2.10, 3.74)	3.13e-12
401	Hypertensive disease	rs653178_C	0.48	119,762	23,634	1.06 (1.04, 1.09)	1.68e-08
401.1	Essential hypertension	rs653178_C	0.48	119,688	23,560	1.06 (1.04, 1.09)	2.00e-08
411.4	Coronary atherosclerosis	rs653178_C	0.48	119,460	9,526	1.09 (1.05, 1.12)	1.27e-07
411	Ischaemic heart disease	rs653178_C	0.48	119,401	9,467	1.09 (1.05, 1.12)	1.33e-07
211	Benign neoplasm of digestive system	rs11264341_C	0.43	117,030	1,504	0.83 (0.77, 0.89)	2.41e-07
274.1	Gout	rs1260326_T	0.39	119,555	1,003	1.26 (1.15, 1.38)	3.86e-07
459.9	Circulatory disease	rs653178_C	0.48	119,677	39,142	1.05 (1.03, 1.06)	2.24e-06
411.2	Myocardial infarction	rs653178_C	0.48	113,559	3,625	1.12 (1.07, 1.18)	2.80e-06
557.1	Coeliac disease	rs1165151_G	0.45	99,783	549	1.33 (1.18, 1.51)	4.30e-06
557.1	Coeliac disease	rs653178_C	0.48	99,965	550	1.31 (1.16, 1.48)	9.28e-06
427.2	Atrial fibrillation and flutter	rs6598541_A	0.35	113,261	4,333	1.11 (1.06, 1.16)	9.92e-06
960	Poisoning by antibiotics	rs1165151_G	0.45	112,343	1,027	0.82 (0.75, 0.90)	1.22e-05
535	Gastritis and duodenitis	rs478607_G	0.15	115,386	5,233	1.12 (1.07, 1.19)	1.34e-05
411.3	Angina pectoris	rs653178_C	0.48	114,967	5,033	1.09 (1.05, 1.14)	3.01e-05
669	Complications of labour and delivery	rs729761_G	0.28	113,240	2,376	1.17 (1.09, 1.26)	3.78e-05

Phecode	Description	SNP_risk allele†	allele_freq	n_total	n_cases	OR (95%CI)	P*
272.11	Hypercholesterolaemia	rs1260326_T	0.39	118,921	10,201	1.07 (1.03, 1.10)	3.82e-05
366	Cataract	rs6770152_G	0.43	116,218	4,567	1.09 (1.05, 1.14)	4.14e-05
471	Nasal polyps	rs10821905_A	0.17	112,745	983	1.26 (1.13, 1.40)	4.61e-05
454.1	Varicose veins of lower extremity	rs2231142_T	0.11	111,390	3,204	0.84 (0.78, 0.92)	5.79e-05
401	Hypertensive disease	rs2079742_T	0.13	115,659	22,832	1.07 (1.03, 1.10)	7.00e-05
401.1	Essential hypertension	rs2079742_T	0.13	115,588	22,761	1.07 (1.03, 1.10)	7.02e-05

*Significance threshold of $p<8.57\times10^{-5}$corresponds to a FDR of $q<0.05$ after correcting the multiple testing.
† Effect allele was harmonised to the SUA-raising allele defined by Köttgen *et al*.(151)

Table 5 - 3: PheWAS associations assessed by conventional MR IVW and MR Egger analysis.

Disease outcomes	**MR IVW**			**MR Egger**			
	OR (95%CI)	**P** *effect*	**Power***	**OR (95%CI)**	**P** *effect*	**P** *pleiotropy*	**Power***
Gout	4.88 (3.91, 6.09)	3.55e-15	1.00	4.58 (2.72, 7.72)	1.76e-06	0.73	1.00
Inflammatory polyarthropathies‡	1.22 (1.11, 1.34)	1.10e-04	0.99	1.15 (1.01, 1.31)	0.03	0.23	0.83
Essential hypertension	1.08 (1.03, 1.14)	5.07e-03	0.82	0.93 (0.83, 1.05)	0.23	1.13e-03	0.73
Hypertensive disease	1.08 (1.03, 1.14)	4.23e-03	0.82	0.93 (0.83, 1.05)	0.24	1.19e-03	0.73
Myocardial infarction	1.16 (1.03, 1.30)	0.02	0.70	1.03 (0.84, 1.27)	0.75	0.13	0.08
Coeliac disease	1.41 (1.05, 1.89)	0.02	0.72	1.31 (0.68, 2.54)	0.41	0.75	0.48
Disorders of metabolism‡	1.07 (1.01, 1.14)	0.03	0.52	1.01 (0.91, 1.14)	0.80	0.18	0.06
Coronary atherosclerosis	1.07 (0.99, 1.15)	0.08	0.41	0.99 (0.85, 1.17)	0.95	0.20	0.06
Ischaemic heart disease	1.07 (0.99, 1.15)	0.09	0.41	0.99 (0.85, 1.16)	0.91	0.20	0.06
Angina pectoris	1.04 (0.94, 1.15)	0.41	0.11	0.95 (0.80, 1.12)	0.51	0.11	0.15
Atrial fibrillation and flutter	1.01 (0.91, 1.12)	0.87	0.05	0.90 (0.75, 1.08)	0.23	0.07	0.41
Circulatory disease	1.04 (1.00, 1.09)	0.08	0.40	0.97 (0.89, 1.07)	0.57	0.05	0.26
Varicose veins of lower extremity	0.86 (0.72, 1.02)	0.09	0.55	0.86 (0.67, 1.10)	0.24	0.97	0.55
Disorders of iron metabolism	1.19 (0.74, 1.90)	0.45	0.11	0.79 (0.15, 4.07)	0.77	0.47	0.12
Hypercholesterolaemia	1.14 (0.96, 1.36)	0.12	0.94	1.18 (0.88, 1.58)	0.27	0.78	0.99
Hypothyroidism	1.10 (0.99, 1.23)	0.07	0.39	0.99 (0.75, 1.32)	0.97	0.30	0.05
Disorders of thyroid	1.08 (0.98, 1.20)	0.10	0.31	1.01 (0.79, 1.29)	0.94	0.41	0.05
Benign neoplasm of digestive system	0.93 (0.78, 1.10)	0.36	0.11	0.90 (0.64, 1.26)	0.52	0.79	0.18
Gastritis and duodenitis	0.97 (0.88, 1.07)	0.53	0.09	0.95 (0.80, 1.13)	0.55	0.70	0.16
Nasal polyps	1.08 (0.88, 1.34)	0.45	0.10	1.09 (0.73, 1.60)	0.67	0.98	0.12

Disease outcomes	**MR IVW**			**MR Egger**			
	OR (95%CI)	**P** *effect*	**Power***	**OR (95%CI)**	**P** *effect*	**P** *pleiotropy*	**Power***
Cataract	0.99 (0.90, 1.09)	0.85	0.05	0.91 (0.75, 1.10)	0.34	0.23	0.36
Poisoning by antibiotics	0.85 (0.70, 1.04)	0.14	0.25	1.00 (0.68, 1.48)	1.00	0.28	0.05
Complications of labour and delivery‡	0.89 (0.76, 1.03)	0.12	0.30	0.78 (0.59, 1.02)	0.08	0.20	0.83
Other diseases of respiratory system‡	1.11 (0.94, 1.31)	0.19	0.22	1.16 (0.92, 1.46)	0.22	0.64	0.42
Disorder of skin and subcutaneous tissue‡	0.99 (0.93, 1.06)	0.77	0.06	0.98 (0.89, 1.09)	0.75	0.85	0.09

‡ Disease outcomes identified from sex-stratified PheWAS analysis.

*The statistical power of MR analyses was calculated by using the non-centrality parameter (NCP) based approach (254); the overall proportion of variance (adjusted R^2) of SUA level explained by the genetic instruments was estimated to be 7.0%.(151)

5.5 Discussion

In PheWAS analysis by using SUA-associated SNPs as genetic instruments, we identified 32 pairs of genotype-phenotype associations, which covered a wide range of phenotypic categories including endocrine/metabolic diseases, cardiovascular diseases, and autoimmune disorders. Our PheWAS analysis replicated 14 pairs of previously known genotype-phenotype (or closely related phenotypic groups) associations reported in the GWAS Catalog (see online **Supplementary Table 5-2** and **Table 5-2**). For example, rs653178 (*ATXN2/SH2B3* locus) was previously reported to be associated with diastolic blood pressure,(475) myocardial infarction (476), peripheral artery disease (477), coeliac disease (478), and serum thyroid peroxidase antibody levels (479). In our PheWAS, this SNP was statistically significantly associated with the same phenotypes (i.e. coeliac disease, myocardial infarction) or similar phenotypic groups (i.e. hypertension, circulatory and heart diseases, hypothyroidism and other disorders of thyroid). Our study also replicated the findings of the largest GWAS performed by Köttgen and the findings of the most recent candidate gene-based association study conducted in UK Biobank, which indicated that 2 SUA-related SNPs (rs12498742 at the *SLC2A9* locus and rs2231142 at the *ABCG2* locus) are statistically significantly associated with gout at GWAS p-value threshold ($p<5.0\times10^{-8}$) (151, 480). We also identified 18 novel genotype-phenotype associations (at the PheWAS threshold of $p<8.57\times10^{-5}$), of which the association between rs1165151 (*SLC17A3*) and disorders of iron metabolism had the smallest p-value ($p=1.23\times10^{-19}$).

We performed conventional MR analysis, using the IVW method, to investigate whether there was a potentially causal link between SUA level and the 25 unique disease groups/outcomes identified from PheWAS. The results of MR IVW analysis suggested a potential causal effect of SUA level on 3 disease groups including inflammatory polyarthropathies, hypertensive disease, disorders of metabolism and 4 specific individual disease outcomes including gout, essential hypertension, myocardial infarction, coeliac disease. When adopting the advanced MR Egger analysis to account for potential pleiotropic effects, it is indicated that, except for gout and its umbrella disease group, inflammatory polyarthropathies, all the other putative causal associations suggested by MR IVW analysis were probably inflated by the presence of pleiotropy. However, although the MR Egger is more robust in dealing with pleiotropy, this method is not infallible (481). Intuitively, the genetic instrument with larger effect on SUA level is expected to have a larger effect on disease outcome and would exert stronger influence in the MR Egger regression model. With in-depth examination of the individual SNP effects on SUA level against the SNP effects on

disease outcomes (see online **Supplementary Figures 5-5**, **5-6**, **5-7**, and **5-8**), we found that the outlying variant (rs12498742 at *SLC2A9*) that had the strongest association with SUA level showed a negative (null) effect on essential hypertension and hypertensive disease, which reversed the sign of the overall putative causal effect and led to a rejection of the intercept test. Given the influence of the outlying variant, the unbalanced pleiotropy and relatively small statistical power (power=0.73), we would interpret that unbalanced pleiotropy between SUA level and hypertension is an issue for their causal inference in MR Egger analysis.

Previous observational studies have reported a sex difference in the association between SUA level and the development of cardiovascular diseases (235, 482-485), but few studies have addressed the sex difference by using a MR approach to remove the influence of environmental confounders. Our study identified a few more cardiovascular diseases (e.g. coronary atherosclerosis, ischaemic heart disease) that were potentially causally linked with the genetic variation of SUA level in women, but not in men. These MR findings were concordant with results from observational studies, which indicated that the relationship between SUA level and cardiovascular disease was particularly strong in women, especially for heart disease (235, 486, 487). Although these putative causal associations specific to females were not verified by MR Egger, this may be due to the decreased statistical power of MR Egger (and a higher risk of type 2 error). The biological mechanism that could lead the association of SUA level with cardiovascular disease to be more pronounced in women than in men remains a matter for further investigation.

To gain a further exploratory sense of pleiotropy in PheWAS findings, we applied the HEIDI test to exclude the PheWAS associations that were probably caused by genetic linkage. The HEIDI test indicated that several PheWAS associations were likely driven by LD. For instance, the outstanding PheWAS association between disorders of iron metabolism and the SNP rs1165151 at the *SLC17A3* locus was not consistent with a pleiotropic model and further examination found the SUA-associated SNP rs1165151 was located in linkage disequilibrium (LD) (r^2=0.24) with the rs17342717 variant at the *SLC17A1* locus, which was strongly associated with disorders of iron metabolism ($p=1.69\times10^{-129}$). This SNP (rs17342717) is also associated with red blood cell traits and serum iron levels in previous GWAS (488, 489). We suggest that the implications of these findings have wider relevance for PheWAS studies. Typically, associations of a single SNP with multiple phenotypes were claimed to be due to pleiotropy in previous PheWAS (490, 491). However, as PheWAS

focused on single variant without considering the correlations between SNPs, we would suggest that an additional examination of LD is necessary when we identify pleiotropic links.

In contrast, the pattern of shared regional genetic associations of SUA level with multiple disease outcomes at *ATXN2/S2HB3* locus was more consistent with a pleiotropic model, where we interpreted this locus influenced a cluster of cardiovascular diseases and autoimmune disorders. However within the *ATXN2/S2HB3* locus there are three leading SNPs (rs653178, rs4766578, and rs3184504) in high LD (r^2 =0.99). In this case, the HEIDI test was unable to provide an indication of whether the observed associations are due to pleiotropy or genetic linkage, as it was difficult to infer the causal variant. Although SNP rs653178 was reported as the lead variant influencing SUA level at this locus in GWAS, the potential biological mechanism underlying this effect is unclear (151). Furthermore, although, an effect of rs653178 on the regulation of blood pressure, cardiovascular diseases and coeliac disease has been suggested by a few GWAS (475-478), a clear biological explanation for this role could not be demonstrated. Evidence from the functional follow-up of the *S2HB3* gene indicated that rs3184504 may be the causal variant, as the *S2HB3* gene encodes one of the S2HB family proteins, which have a diverse physiological roles on haematopoiesis, immune response and signalling, and variation in rs3184504 may introduce a new phosphorylation site affecting the function of the S2HB protein (492, 493). We believe that further uncovering of the biological functions of this pleiotropic locus (e.g. gene function follow-up, expression quantitative trait loci [eQTL] analysis) might be helpful to understand the complex underlying relationship of SUA level with cardiovascular and autoimmune diseases.

To judge the nature of PheWAS associations more comprehensively, it is important to consider different lines of evidence, including underlying biological plausibility. Therefore, we would like to highlight a few disease outcomes beyond joints and cardiovascular events. The sex-stratified MR IVW analysis identified that unspecified diseases in respiratory system were potentially causally linked with SUA level in women (with the MR Egger analysis showing a consistent causal effect). This finding is consistent with recently published experimental studies, which demonstrated that human airway epithelial cells and lung tissue expressed a functional UA production/secretion system and UA was crucial in mediating the development of allergic airway diseases and regulating the antigen-specific T-cell proliferation (494-497). It was also speculated that fine, inhaled particulate matter (PM) can induce increased UA production in the human airway which may contribute to allergic sensitisation and asthma pathogenesis (498). Evidence from epidemiological studies in

relation to the association between SUA level and respiratory diseases has not been well explored (499, 500). Our findings support further investigation of the clinical relevance of SUA level in lung health and respiratory diseases.

Key strengths of our study included its potential to make novel discoveries in genotype-phenotype associations and to identify novel cross-phenotype associations, possibly reflecting common aetiology or causal mechanisms. Unlike the genome, for which genetic structure can be measured by reliable biological techniques, the definition of the phenome varies across studies. Current published PheWAS have been limited primarily to billing ICD-9-CM to phecode system, and the method for aggregating ICD-9-CM codes into phecodes has proven to be valuable in previous PheWAS studies (433, 459). Our work broadened the utility of the PheCODE system and illustrated the process of adopting the system to the updated ICD-10 version to define the phenome framework. Our mapping process revealed some potential shortcomings of the current PheCODE system (e.g. the ICD-10 code involving the personal or family history were missing elements in the PheCODE system), which should be improved as a future undertaking. Recent methodological applications (e.g. tree-structured phenotypic model [TreeWAS]) can be applied for future PheWAS analyses (501). As we were preparing the manuscript for submission, a web-resource within UK Biobank, the GeneATLAS, was released in the bioRxiv (prior to peer review) (502). We checked our PheWAS findings in this database, but only 10 of the 31 SUA related SNPs were included in their database (and associations with some disease outcomes were replicated for these SNPs) (502). We focused on the causal relationships between SUA level and binary disease outcomes in MR analyses, and these findings were complementary to MR estimates of urate archived in the MR-Base database (http://eve.mrbase.org/), which mainly focused on quantitative traits.

On the other hand, our analysis was limited to phenotypes with no less than 200 cases, therefore diseases with relatively low prevalence were not analysed. As the UK biobank grows we expect to perform PheWAS and MR analysis for more phenotypes, with the priority given for the ones of which the relationships with SUA level are much controversial such as dementia (503, 504). Furthermore for some analysed phenotypes, our PheWAS analysis may still have low power to detect small effect size. The use of the interim release of UK Biobank data and focusing on a very homogenous population (self-reported British confirmed by PCAs) limited the power of this study. Additionally, we did not analyse the self-reported UK Biobank data to avoid information bias, but this may impact the comprehensiveness of PheWAS and reduce the precision of MR estimates. To remedy this

limitation, we performed a sensitivity analysis for gout by comparing the MR estimates for hospital-diagnosed gout, self-reported gout and hospital-diagnosed/self-reported gout (see online **Supplementary Table 5-8**). The MR estimates are consistently statistically significant in any of the cases but with differences in their effect sizes. These differences might be due to the fact that gout cases ascertained from hospital discharge coding may be unrepresentative of gout, given hospitalised gout is more likely to be complicated by co-morbidities, as reported by Cadzow *et al* (480). While making efforts to dissect the PheWAS associations with different models, given the complexity of human genetic structure, these models are not mutually exclusive and each model has its own methodological limitations, thus strong conclusions are not always possible. Therefore, the realistic goal for the present study was to assess different lines of evidence (i.e. causality, pleiotropy or genetic linkage) in order to characterize the identified PheWAS associations in relation to SUA level. It would be beneficial to assess whether measured SUA level, rather than its genetic proxy, are also associated with the observed disease outcomes, but data on the SUA biomarker are not yet available in UK Biobank.

Overall, this PheWAS analysis demonstrated that SUA level shares genetic risk loci with multiple disease outcomes, particularly cardiovascular/metabolic diseases and autoimmune disorders. These findings provide the rationale for further investigation of whether these associations are causal. Our study indicated a putative causal effect of SUA level on 3 disease groups (inflammatory polyarthropathies, hypertensive disease, and disorders of metabolism) and 4 specific disease outcomes (gout, essential hypertension, coeliac disease, and myocardial infarction); when balancing out the pleiotropy, a robust conclusion about causality was made for gout and its encompassing disease group, inflammatory polyarthropathies. Unbalanced pleiotropy was identified as an issue for the causal inference on the association between SUA level and hypertension. Other potential causal relevance of SUA level with respiratory diseases and ocular abnormalities are also worthy of further investigation. When interpreting the PheWAS associations from a view of pleiotropy, our analysis highlighted a key pleiotropic locus that influenced SUA level and multiple cardiovascular and autoimmune diseases. A further functional annotation of this locus might be helpful to understand the biological pathways that contribute to the phenotypic associations between SUA level and cardiovascular diseases (including hypertension).

Acknowledgements: This research has been conducted using the UK Biobank resource (approval number: 10775). We wish to kindly thank all participants in UK Biobank.

Contributors: E.T. and H.C. conceived the study and X.L. contributed to the study design. X.L. performed the data analysis. X.L., X.M., W.Q., A.G., J.C.D. and T.V. contributed to the mapping of ICD-10/9 code to phecode. X.L. wrote the manuscript. All authors critically reviewed the manuscript and contributed important intellectual content. All authors have read and approved the final manuscript as submitted.

Funding: X.L. and X.M. are supported by the China Scholarship Council. E.T. is supported by a CRUK Career Development Fellowship.

Competing interests: None declared.

Ethics approval: UK Biobank has approval from the North West Multi-Centre Research Ethics Committee (11/NW/0382) and obtained written informed consent from all participants prior to the study. This study did not need to re-contact the participants and no separate ethics approval was required according to the Ethics and Governance Framework (EGF) of UK Biobank.

Provenance and peer review: Not commissioned; externally peer reviewed.

Data sharing statement: All the data generated or analysed during this study are included in this published article and in Supplementary Information files. Further inquiry regarding to data availability, analysis methods and results would be addressed to X.L. (xue.li@ed.ac.uk).

5.6 Supplementary information

Online Supplementary text

Phenotyping and mapping ICD-10/9 to phecode

ICD-9/10 codes were organised in a hierarchical tree-like structure. The individual ICD-9/10 codes could not be directly used for PheWAS analysis, because they were not designed for representing distinct disease phenotypes. To aggregate the ICD codes, the PheCODE schema has been successfully adopted in a number of PheWAS to combine one or more individual ICD codes into distinct phenotype groups (433, 459). However, since the current version of the PheCODE was developed based on ICD-9-Clinical Modification (CM), the phencode algorithm was not directly applicable to the ICD-10 coding system in the UK Biobank. To develop an aggregation method for PheWAS analysis in UK Biobank, we collaborated with the Electronic Medical Records and Genomics (eMERGE) group of Vanderbilt University Medical Center and mapped ICD-10 codes to phecodes in both direct and indirect ways. We mapped the ICD-10 codes to phecodes directly if their descriptions matched each other regardless of capitalisation. Otherwise, we used the unified medical language system (UMLS) to map the ICD-10 code to ICD-9-CM (or map the ICD-10 code to systematised nomenclature of medicine clinical terms [SNOMED CT] code first and then to ICD-9-CM) and then used the previous mapping of ICD-9-CM to phecode to finally link the ICD-10 to phecode. The ICD-9 codes in UK Biobank were directly mapped to the phecodes through the first fourth or full (five) digital codes or through the descriptions regardless of capitalisation.

MR IVW, MR Egger and HEIDI test

MR IVW For each independent genetic instrument i, the causal effect of SUA level on the disease outcome (denoted as $b_{xy(i)}$) was estimated by the ratio method, in which the coefficient from the regression of outcome on the genetic variant (using individual-level data from the UK Biobank and denoted as $b_{zy(i)}$) was divided by the coefficient from the regression of SUA level on the genetic variant (using the summary-level GWAS data made available by Köttgen *et al* and denoted as $b_{zx(i)}$) (151). The overall causal effect of SUA level on the outcome mediated by all 31 genetic instruments was estimated by pooling the individual effect estimates of each SNP using the IVW method (474).

MR Egger Briefly, instead of assuming that the genetic instruments are only associated with SUA level (no pleiotropy criterion of MR), the MR Egger uses a weighted linear regression

to regress the effect estimates of SNP-outcome associations against the effect estimates of SNP-SUA associations with the intercept unconstrained. The unconstrained intercept represents the average pleiotropic effects across the genetic variants (with a zero intercept indicating that there are no direct pleiotropic effects or the pleiotropic effects are balanced among the multiple genetic instruments). The slope coefficient from the MR Egger regression represents the overall estimate of the causal effect after accounting for the pleiotropic effects of multiple genetic instruments (469).

HEIDI test The HEIDI test was firstly proposed by Zhu *et al* (470), but the principle of this test can be broadly applied to any pair of traits. The rationale and mathematical theories of this metric are explained elsewhere in detail (470). The HEIDI method assumes only one causal variant affected both the SUA level and disease outcome (via either vertical pleiotropy [including causality] or horizontal pleiotropy) within a genetic region. If we denote the b_{zx} as the effect estimate of a genetic variant on SUA level, and b_{zy} as the effect estimate of a genetic variant on disease outcome, the effect estimate of SUA level on disease outcome mediated by the genetic component could be calculated by the ratio method:

$$b_{xy} = \frac{b_{zy}}{b_{zx}}$$

If we describe the casual variant as SNP_0, under the Hardy-Weinberg equilibrium, for any SNP_i in LD with the causal variant, the effect estimate $b_{xy(i)}$ calculated by the ratio method should be identical to $b_{xy(0)}$:

$$b_{xy(i)} = \frac{b_{zy(i)}}{b_{zx(i)}} = \frac{b_{zy(0)}\, r_{0i}\, \sqrt{\mathrm{h}_0/h_i}}{b_{zx(0)}\, r_{0i}\, \sqrt{\mathrm{h}_0/h_i}} = \frac{b_{zy(0)}}{b_{zx(0)}} = b_{xy(0)}$$

where r_{0i} is the LD correlation between the casual variant SNP (0) and SNP (i), and $h_{0/i}$ is determined by the allele frequency ($h = 2p(1-p)$). Thus testing linkage against pleiotropy is equivalent to testing if there were any heterogeneity between $b_{xy(0)}$ and $b_{xy(i)}$. If we defined

$$d_i = b_{xy(i)} - b_{xy(0)}$$

then it is equivalent to testing if $d_i = 0$.

With the matrix of any pair of SNPs(i, j) , the covariance could be calculated by

$$cov\left(b_{xy(i)}, b_{xy(j)}\right) = \frac{r_{ij}}{b_{zx(i)}b_{zx(j)}}\sqrt{var(b_{zy(i)})var(b_{zy(j)})} + b_{xy(i)}b_{xy(j)}\left(\frac{r_{ij}}{z_{zx(i)}z_{zx(j)}} - \frac{1}{z_{zx(i)}^2\ z_{zx(j)}^2}\right)$$

$$cov\left(d_i, d_j\right) = cov\left(b_{xy(i)}, b_{xy(j)}\right) - cov\left(b_{xy(i)}, b_{xy(0)}\right) - cov\left(b_{xy(j)}, b_{xy(0)}\right) + var(b_{xy(0)})$$

Then, we calculated the Z values of d_i

$$z_{d(i)} = d_i/\sqrt{var(d_i)}$$

Under the null hypothesis, where $d_i = 0$, we have a vector of z_d value that follows the multivariate normal distribution (approximated by the Satterthwaite method) with z_d ~ MVN (0, R), where R is the correlation matrix with the *ij*th element

$$r\left(z_{d(i)}, z_{d(j)}\right) = cov(d_i, d_j)/\sqrt{var(d_i)var(d_j)}$$

The HEIDI statistics was calculated as

$$T_{HEIDI} = \sum_i^m z_{d_{(i)}}^2$$

with m being the number of SNPs associated with SUA level with $p<1.57\times10^{-3}$ (equivalent to $\chi^2>10$).

In the analysis, we defined a region of ±250kb (upstream and downstream) around the locus associated with SUA level. For each locus, we calculated the HEIDI statistic only including SNPs that were associated with SUA level at $p<1.57\times10^{-3}$ (equivalent to $\chi^2>10$) in order to avoid very weak instruments and to increase the power. The larger the heterogeneity, the smaller the HEIDI's P-value, and the higher the probability of association is caused by LD. The pattern of regional genetic association with SUA level and disease outcome was visualised by the LocusZoom (505).

Online Supplementary tables

Supplementary Table 5-1: Sample quality control (QC) of genotype data in UK Biobank.

Supplementary Table 5-2: Summary of the 31 SUA-associated SNPs identified from GWAS.

Supplementary Table 5-3: Demographic characteristics of the sampled UK Biobank participants (n=120,091).

Supplementary Table 5-4: Genotype-phenotype associations identified from PheWAS analysis without adjustment for BMI.

Supplementary Table 5-5: Genotype-phenotype associations identified from sex-stratified PheWAS after correcting multiple testing by FDR.

Supplementary Table 5-6: The mappings of ICD codes to phecodes for the 25 disease outcomes identified from PheWAS.

Supplementary Table 5-7: Sex-stratified MR IVW analysis and MR Egger analysis.

Supplementary Table 5-8: A sensitivity analysis of MR for gout cases defined by multiple criteria.

Supplementary Table 5 - 1: Sample quality control (QC) of genotype data in UK Biobank.

Sample QC variables			**Sample selection**
f.22050	Pass or not	• This variable indicates the genotype quality for ~50,000 samples genotyped by UKBiLEVE array • Sample quality control undertaken by Affymetrix was a DNA quality filter (dish quality control < 0.82) and initial clustering call rate (<97%); • 520 duplicates passed Affymetrix quality control in both samples, 33 passed Affymetrix quality control in 1 sample, and 0 duplicates failed Affymetrix quality control in both samples.	Pass
f.22051	Pass or not	• This variable indicates QC steps performed for ~50,000 samples genotyped by UKBiLEVE array • Sex mismatch (remove sample if submitted gender is different to gender inferred from sex chromosomes); • Final call rate (<95%), • Heterozygosity outliers (outlier if >3sd from mean heterozygosity); • Unintended duplicate (>98% of alleles shared identical by descent); • Ancestry principal components outlier (outlier if >10sd from mean on first 10 principal components).	Pass
f.22010	Pass or not	• This variable indicates the genotype quality for all samples; • Samples showed signs of insufficient data quality; • Individuals with high missing or for which heterozygosity rates were not explained by the long runs of homozygosity (ROH) nor mixed ethnicity.	Pass
Population structure			
f.22006	Genetic ethnic group	• Ethnic groups determined by principle genetic components.	Select self-reported British and confirmed to be Europeans by PCAs
f.21000	Ethnic background	• Self-reported ethnic background	
f.22009	Genetic principal components (PCAs)	• The first fifteen PCAs were available for the samples	
Sex mismatch			
f.22001	Genetic sex	• Sex inferred from X chromosome genotypes	Remove samples with sex mismatch
f.31	Self-reported sex	• Sex self-reported from the baseline assessment	

Supplementary Table 5 - 2: Summary of the 31 SUA-associated SNPs identified from GWAS.

Genetic predictors				Overall effect			Male-specific effect			Female-specific effect			Other associated phenotypes reported in the human GWAS catalog
SNPs	Chr	Effect allele	Closest/GRAIL gene	Beta	SE	p_val	Beta	SE	p_val	Beta	SE	p_val	
rs10480300	7	T	*PRKAG2/PRKAG2*	0.032	0.006	9.37E-07	0.043	0.010	1.70E-05	0.024	0.008	3.20E-03	Red blood cell traits(506)
rs10821905	10	A	*A1CF/ASAH2*	0.053	0.007	3.45E-12	0.042	0.011	3.80E-04	0.060	0.009	2.50E-10	None
rs11264341	1	C	*TRIM46/PKLR*	0.048	0.006	1.04E-14	0.055	0.009	1.10E-08	0.044	0.007	9.10E-09	None
rs1165151	6	G	*SLC17A1/SLC17A3*	0.092	0.005	4.52E-60	0.096	0.008	1.30E-28	0.089	0.007	4.20E-37	None
rs1171614	10	C	*SLC16A9/SLC16A9*	0.074	0.007	6.48E-23	0.086	0.011	1.90E-13	0.067	0.009	3.00E-13	Acylcarnitine levels,(507) Glycerophospholipid levels,(507) Blood metabolite levels(508)
rs1178977	7	A	*BAZ1B/MLXIPL*	0.050	0.007	6.68E-12	0.055	0.011	8.20E-07	0.046	0.009	2.60E-07	None
rs12498742	4	A	*SLC2A9/SLC2A9*	0.380	0.006	0.00E+00	0.269	0.010	6.40E-153	0.460	0.008	0.00E+00	Gout(151)
rs1260326	2	T	*GCKR/GCKR*	0.077	0.006	1.31E-40	0.091	0.009	3.00E-25	0.063	0.007	1.90E-04	Gout,(509) C-reactive protein levels,(510) Dyslipidaemia,(511) Haematological and biochemical traits,(512) Chronic kidney disease,(513) Hypertriglyceridaemia(514)
rs1394125	15	A	*UBE2Q2/NRG4*	0.043	0.006	9.78E-11	0.060	0.010	5.50E-06	0.032	0.008	1.00E-04	Chronic kidney disease(513), Kidney function(515)
rs1471633	1	A	*PDZK1/PDZK1*	0.061	0.005	1.40E-26	0.069	0.008	3.50E-15	0.054	0.007	1.60E-14	None
rs164009	17	A	*QRICH2/PRPSAP1*	0.029	0.006	7.06E-07	0.024	0.009	6.20E-03	0.032	0.007	8.20E-06	None
rs17050272	2	A	*INHBB/INHBB*	0.037	0.006	9.36E-09	0.049	0.010	6.50E-07	0.030	0.008	1.90E-04	Glomerular filtration rate(creatinine)(515)
rs17632159	5	G	*TMEM171/TMEM171*	0.038	0.006	2.00E-09	0.043	0.010	1.30E-05	0.039	0.008	1.10E-06	None
rs17786744	8	G	*STC1/STC1*	0.031	0.005	8.82E-08	0.033	0.009	2.10E-04	0.029	0.007	2.10E-04	None
rs2078267	11	C	*SLC22A11/SLC22A11*	0.078	0.006	8.73E-36	0.085	0.009	2.90E-19	0.071	0.007	5.70E-20	Gout,(333) Cardiovascular disease risk factors(333)
rs2079742	17	T	*BCAS3/C17orf82*	0.051	0.008	6.24E-09	0.054	0.013	5.60E-05	0.048	0.010	1.00E-05	Metabolite levels (small molecules and protein measures)(516)

Genetic predictors				Overall effect			Male-specific effect			Female-specific effect			Other associated phenotypes reported in the human GWAS catalog
SNPs	**Chr**	**Effect allele**	**Closest/GRAIL gene**	**Beta**	**SE**	**p_val**	**Beta**	**SE**	**p_val**	**Beta**	**SE**	**p_val**	
rs2231142	4	T	*ABCG2/ABCG2*	0.220	0.009	4.43E-116	0.270	0.014	3.80E-75	0.181	0.011	1.30E-52	Gout(151)
rs2307394	2	C	*ORC4L/ACVR2A*	0.035	0.006	7.26E-09	0.036	0.009	1.20E-04	0.034	0.007	4.70E-06	None
rs2941484	8	T	*HNF4G/HNF4G*	0.049	0.006	3.91E-17	0.048	0.009	6.20E-08	0.046	0.007	1.30E-10	None
rs3741414	12	C	*INHBC/INHBE*	0.071	0.007	9.79E-22	0.091	0.011	7.00E-16	0.057	0.009	4.30E-10	None
rs478607	11	G	*NRXN2/SLC22A12*	0.048	0.007	5.31E-10	0.058	0.012	9.60E-07	0.043	0.009	8.80E-06	None
rs642803	11	C	*OVOL1/LTBP3*	0.043	0.005	4.51E-14	0.047	0.008	8.00E-08	0.042	0.007	2.10E-09	None
rs653178	12	C	*ATXN2/SH2B3*	0.036	0.005	2.45E-10	0.044	0.009	7.50E-07	0.032	0.007	5.50E-06	Coeliac disease,(517) Diastolic Blood Pressure,(475) Chronic kidney disease,(513) Serum thyroid peroxidase antibody levels,(479) Rheumatoid arthritis,(518) Peripheral artery disease,(477) Myocardial infarction,(519) Inflammatory bowel disease(520)
rs6598541	15	A	*IGF1R/IGF1R*	0.044	0.006	5.20E-13	0.039	0.009	2.70E-05	0.050	0.007	1.60E-11	None
rs675209	6	T	*RREB1/RREB1*	0.063	0.006	1.38E-21	0.060	0.010	3.30E-09	0.064	0.008	2.00E-15	Gout,(333) Cardiovascular disease risk factors(333)
rs6770152	3	G	*SFMBT1/MUSTN1*	0.048	0.006	2.66E-16	0.052	0.009	6.70E-09	0.047	0.007	6.00E-11	None
rs7188445	16	G	*MAF/MAF*	0.032	0.006	1.15E-07	0.025	0.009	7.90E-03	0.040	0.007	6.40E-08	None
rs7193778	16	C	*NFAT5/NFAT5*	0.047	0.008	2.36E-08	0.048	0.012	2.10E-04	0.045	0.010	1.00E-05	None
rs7224610	17	C	*HLF/HLF*	0.038	0.006	4.74E-11	0.043	0.009	9.00E-07	0.034	0.007	3.00E-06	None
rs729761	6	G	*VEGFA/VEGFA*	0.046	0.006	3.05E-12	0.047	0.010	3.20E-06	0.047	0.008	3.20E-06	None
rs742132	6	A	*LRRC16A/LRRC16A*	0.035	0.006	1.90E-08	0.035	0.006	1.90E-08	0.035	0.006	1.90E-08	Cardiovascular disease risk factors,(333) Haematological and biochemical traits,(512) Metabolite levels(521)

Abbreviations: SNP=Single Nucleotide Polymorphism; GWAS= Genome Wide Association Study; GRAIL=Gene Recognition and Analysis Internet Link.

Supplementary Table 5 - 3: Demographic characteristics of the sampled UK Biobank participants (n=120,091).

Continuous variable	**Mean (S.D.)**	**Number of missing**
Age[#]	64.86 (7.95) years	0
Standing height	168.79 (9.20) cm	200
Weight	78.68 (16.08) kg	288
BMI[†]	27.54 (4.83) kg/m^2	329
Categorical variable	**Levels**	**Number of participants**
Sex	Male	56,845
	Female	63,246
Assessment center	Barts	2,004
	Birmingham	5,427
	Bristol	10,808
	Bury	7,751
	Cardiff	4,539
	Croydon	4,933
	Edinburgh	4,257
	Glasgow	4,783
	Hounslow	5,082
	Leeds	11,188
	Liverpool	8,404
	Manchester	3,278
	Middlesborough	5,563
	Newcastle	9,592
	Nottingham	8,429
	Oxford	3,163
	Reading	7,021
	Sheffield	7,779
	Stockport(pilot)	88
	Stoke	5,229
	Swansea	571
	Wrexham	202

[#] Variable represents the mean age of participants at the year when we assessed their medical records.
[†]BMI (body mass index) was calculated as the weight divided by the square of the height

Supplementary Table 5 - 4: Genotype-phenotype associations identified from PheWAS analysis without adjustment for BMI.

Phecode	Description	SNP_risk allele	n_total	n_cases	allele_freq	HWE_p	beta	SE	OR (95%CI)	P
274.1	Gout	rs2231142_T	119884	1008	0.113	0.642	0.619	0.057	1.86 (1.66, 2.08)	4.17E-27
275.1	Disorders of iron metabolism	rs1165151_G	119389	205	0.451	0.920	1.270	0.127	3.56 (2.78, 4.57)	1.34E-23
244.4	Hypothyroidism NOS	rs653178_C	119145	4169	0.482	0.690	0.184	0.023	1.20 (1.15, 1.26)	3.19E-16
246	Other disorders of thyroid	rs653178_C	119929	4953	0.483	0.708	0.163	0.021	1.18 (1.13, 1.23)	4.56E-15
274.1	Gout	rs12498742_A	119284	1007	0.233	0.773	0.429	0.060	1.54 (1.37, 1.73)	1.00E-12
275.1	Disorders of iron metabolism	rs742132_A	119598	205	0.293	0.790	1.029	0.148	2.80 (2.10, 3.74)	3.15E-12
211	Benign neoplasm of other parts of digestive system	rs11264341_C	117350	1507	0.429	0.949	-0.189	0.037	0.83 (0.77, 0.89)	2.68E-07
411.8	Other chronic ischaemic heart disease	rs653178_C	119699	9495	0.483	0.720	0.075	0.016	1.08 (1.05, 1.11)	1.24E-06
411.4	Coronary atherosclerosis	rs653178_C	119787	9583	0.483	0.716	0.075	0.015	1.08 (1.05, 1.11)	1.41E-06
411	Ischaemic Heart Disease	rs653178_C	119728	9524	0.483	0.723	0.075	0.015	1.08 (1.05, 1.11)	1.47E-06
274.1	Gout	rs1260326_T	119884	1008	0.393	0.813	0.218	0.045	1.24 (1.14, 1.36)	1.48E-06
401	Hypertension	rs653178_C	120091	23755	0.483	0.720	0.049	0.011	1.05 (1.03, 1.07)	3.59E-06
427.2	Atrial fibrillation and flutter	rs6598541_A	113564	4368	0.353	0.549	0.106	0.023	1.11 (1.06, 1.16)	3.69E-06
401.1	Essential hypertension	rs653178_C	120017	23681	0.483	0.707	0.049	0.011	1.05 (1.03, 1.07)	4.20E-06
557.1	Coeliac disease	rs1165151_G	100035	550	0.450	0.904	0.285	0.062	1.33 (1.18, 1.50)	5.11E-06
960	Poisoning by antibiotics	rs1165151_G	112628	1031	0.451	0.864	-0.199	0.045	0.82 (0.75, 0.89)	8.01E-06
557.1	Coeliac disease	rs653178_C	100218	551	0.482	0.306	0.271	0.061	1.31 (1.16, 1.48)	8.27E-06
411.2	Myocardial infarction	rs653178_C	113854	3650	0.482	0.603	0.105	0.024	1.11 (1.06, 1.16)	1.31E-05
535	Gastritis and duodenitis	rs478607_G	115704	5252	0.152	0.914	0.114	0.027	1.12 (1.06, 1.18)	2.50E-05
459.9	Circulatory disease NEC	rs653178_C	120005	39322	0.483	0.702	0.037	0.009	1.04 (1.02, 1.06)	4.54E-05
471	Nasal polyps	rs10821905_A	113054	986	0.172	0.725	0.227	0.056	1.26 (1.12, 1.40)	5.06E-05
366	Cataract	rs6770152_G	116535	4592	0.427	0.063	0.087	0.022	1.09 (1.05, 1.14)	6.44E-05
292.1	Aphasia/speech disturbance	rs164009_A	116338	425	0.385	0.969	0.291	0.074	1.34 (1.16, 1.55)	8.38E-05

Supplementary Table 5 - 5: Genotype-phenotype associations identified from sex-stratified PheWAS after correcting multiple testing by FDR.*

Phecode	Description	SNP_effect _allele	n_total	n_cases	allele_freq	hwe_p	Beta	SE	OR (95%CI)	P
Significant associations identified from PheWAS analysis in men										
274.1	Gout	rs2231142_T	56528	885	0.113	0.939	0.681	0.061	1.98 (1.75, 2.23)	8.23E-29
274.1	Gout	rs12498742_A	56253	884	0.234	0.876	-0.430	0.064	1.54 (1.36, 1.74)	2.68E-11
714	Inflammatory polyarthropathies	rs2231142_T	56306	2685	0.113	0.956	0.209	0.042	1.23 (1.14, 1.34)	5.82E-07
274.1	Gout	rs1260326_T	56528	885	0.393	0.554	0.238	0.049	1.27 (1.15, 1.40)	1.01E-06
401	Hypertensive disease	rs653178_C	56668	13027	0.482	0.834	0.073	0.015	1.08 (1.04, 1.11)	1.34E-06
401.1	Essential hypertension	rs653178_C	56630	12989	0.482	0.824	0.073	0.015	1.08 (1.04, 1.11)	1.42E-06
277	Disorders of metabolism‡	rs3741414_C	56503	8017	0.243	0.604	-0.096	0.021	1.10 (1.06, 1.15)	3.98E-06
714	Inflammatory polyarthropathies‡	rs1260326_T	56306	2685	0.393	0.657	0.120	0.029	1.13 (1.07, 1.19)	3.09E-05
512	Other diseases of respiratory system‡	rs1471633_A	55070	977	0.462	0.930	-0.193	0.046	0.82 (0.75, 0.90)	3.25E-05
274.1	Gout	rs642803_C	55592	872	0.460	0.573	-0.198	0.050	1.22 (1.11, 1.34)	6.17E-05
Significant associations identified from PheWAS analysis in women										
244.4	Hypothyroidism NOS	rs653178_C	62363	3362	0.483	0.030	0.205	0.025	1.23 (1.17, 1.29)	7.95E-16
246	Other disorders of thyroid	rs653178_C	62970	3969	0.483	0.096	0.179	0.023	1.20 (1.14, 1.25)	2.51E-14
211	Benign neoplasm of digestive system	rs11264341_C	61908	896	0.431	0.872	0.257	0.048	0.77 (0.70, 0.85)	7.47E-08
411.4	Coronary atherosclerosis	rs653178_C	62971	2997	0.483	0.185	0.113	0.027	1.12 (1.06, 1.18)	2.71E-05
411	Ischaemic Heart Disease	rs653178_C	62954	2980	0.483	0.157	0.112	0.027	1.12 (1.06, 1.18)	3.35E-05
669	Complications of labor and delivery#	rs729761_G	59622	2376	0.283	0.958	-0.159	0.039	1.17 (1.09, 1.27)	3.78E-05
401.1	Essential hypertension	rs2079742_T	60918	10237	0.136	0.339	-0.098	0.024	1.10 (1.05, 1.16)	5.35E-05
689	Disorder of skin and subcutaneous tissue#	rs2231142_T	63094	6142	0.113	0.349	-0.126	0.031	0.88 (0.83, 0.94)	5.72E-05
401	Hypertensive disease	rs2079742_T	60952	10271	0.136	0.310	-0.098	0.024	1.10 (1.05, 1.16)	5.78E-05
535	Gastritis and duodenitis	rs478607_G	60962	2845	0.153	0.790	0.144	0.036	1.15 (1.08, 1.24)	7.46E-05

*Significance threshold of $p<8.57\times10^{-5}$ corresponds to a FDR of $q<0.05$ after correcting the multiple testing.

‡ Genotype-phenotype associations that were not identified in PheWAS analysis of overall population.

Genotype-phenotype association that were not identified in PheWAS analysis of overall population.

Supplementary Table 5 - 6: The mappings of ICD codes to phecodes for the 25 disease outcomes identified from PheWAS.

Phecode	ICD9/10	Descriptions
274.1	**Gout**	
ICD-9	274	Gout
	274	Gouty arthropathy
	274.1	Gouty nephropathy
	274.8	Gout with other specified manifestations
	274.9	Gout, unspecified
ICD-10	M10	Gout
	M10.0	Idiopathic gout
	M10.1	Lead-induced gout
	M10.2	Drug-induced gout
	M10.3	Gout due to impairment of renal function
	M10.4	Other secondary gout
	M10.9	Gout, unspecified
714	**Inflammatory polyarthropathies**	
ICD-9	714	Rheumatoid arthritis and other inflammatory polyarthropathies
	714	Rheumatoid arthritis
	714.1	Felty's syndrome
	714.2	Other rheumatoid arthritis with visceral or systemic involvement
	714.3	Juvenile chronic polyarthritis
	714.4	Chronic postrheumatic arthropathy
	714.8	Other specified inflammatory polyarthropathies
	714.9	Unspecified inflammatory polyarthropathy
ICD-10	M05	Seropositive rheumatoid arthritis
		M05.0-M05.9
	M06	Other rheumatoid arthritis
		M06.0-M06.9
	M07	Psoriatic and enteropathic arthropathies
		M07.0-M07.6
	M08	Juvenile arthritis
		M08.0-M08.9
	M09	Juvenile arthritis in diseases classified elsewhere
		M09.0-M09.8
	M10	Gout
		M10.0-M10.9
	M11	Other crystal arthropathies
		M11.0-M11.9
	M12	Other specific arthropathies
		M12.0-M12.8

	M13	Other arthritis
		M13.0-M13.9
	M14	Arthropathies in other diseases classified elsewhere
		M14.0-M14.8
401.1	**Essential hypertension**	
ICD-9	401	Essential hypertension
		401.0-401.9
ICD-10	I10	Essential (primary) hypertension
401	**Hypertensive disease**	
ICD-9	401	Essential hypertension
		401.0-401.9
	402	Hypertensive heart disease
		402.0-402.9
	403	Hypertensive chronic kidney disease
		403.0-403.9
	404	Hypertensive heart and chronic kidney disease
		404.0-404.9
	405	Secondary hypertension
		405.0-405.9
ICD-10	I10	Essential (primary) hypertension
	I11	Hypertensive heart disease
		I11.0-I11.9
	I12	Hypertensive renal disease
		I12.0-I12.9
	I13	Hypertensive heart and renal disease
		I13.0-I13.9
	I15	Secondary hypertension
		I15.0-115.9
411.4	**Coronary atherosclerosis**	
ICD-9	411.81	Acute coronary occlusion without myocardial infarction
	414	Other forms of chronic ischaemic heart disease
		414.00-414.07
	414.1	Aneurysm and dissection of heart
		414.10-414.19
	414.2	Chronic total occlusion of coronary artery
	414.3	Coronary atherosclerosis due to lipid rich plaque
	996.03	Due to coronary bypass graft
	V45.81	Aortocoronary bypass status
	V45.82	Percutaneous transluminal coronary angioplasty status
ICD-10	I20	Angina pectoris
		I20.0-I20.9

	I21	Acute myocardial infarction
		I21.0-I21.9
	I22	Subsequent myocardial infarction
		I22.0-I22.9
	I23	Certain current complications following acute myocardial infarction
		I23.0-I23.9
	I24	Other acute ischaemic heart diseases
		I24.0-I24.9
	I25	Chronic ischaemic heart disease
		I25.0-I25.9
	Z95.1	Presence of aortocoronary bypass graft
	Z95.5	Presence of coronary angioplasty implant and graft
411.2	**Myocardial infarction**	
ICD-9	410	Acute myocardial infarction
		410.0-410.9
	411	Postmyocardial infarction syndrome
	412	Old myocardial infarction
	429.7	Certain sequelae of myocardial infarction, not elsewhere classified
		429.71-429.79
ICD-10	I21	Acute myocardial infarction
		I21.0-I21.9
	I22	Subsequent myocardial infarction
		I22.0-I22.9
	I23	Certain current complications following acute myocardial infarction
		I23.0-I23.9
	I24.1	Dressler syndrome
	I25.2	Old myocardial infarction
	I51.0	Cardiac septal defect
	I51.3	Intracardiac thrombosis, not elsewhere classified
427.2	**Atrial fibrillation and flutter**	
ICD-9	427.3	Atrial fibrillation and flutter
		427.31-427.32
ICD-10	I48	Atrial fibrillation and flutter
		I48.0-I48.9
411.3	**Angina pectoris**	
ICD-9	413	Angina pectoris
		413.0-413.9
ICD-10	I20	Angina pectoris
		I20.0-I20.9
411	**Ischaemic heart disease**	
ICD-9	410	Acute myocardial infarction

		410.0-410.9
	411	Other acute and subacute forms of ischaemic heart disease
		411.0-411.8
	412	Old myocardial infarction
	413	Angina pectoris
		413.0-413.9
	414	Other forms of chronic ischaemic heart disease
		414.0-414.9
	429	Ill-defined descriptions and complications of heart disease
	996.03	Due to coronary bypass graft
ICD-10	I20	Angina pectoris
		I20.0-I20.9
	I21	Acute myocardial infarction
		I21.0-I21.9
	I22	Subsequent myocardial infarction
		I22.0-I22.9
	I23	Certain current complications following acute myocardial infarction
		I23.0-I23.9
	I24	Other acute ischaemic heart diseases
		I24.0-I24.9
	I25	Chronic ischaemic heart disease
		I25.0-I25.9
454.1	**Varicose veins of lower extremity**	
ICD-9	454	Varicose veins of lower extremities
		454.0-454.9
ICD-10	I83	Varicose veins of lower extremities
		I83.0-I83.9
459.9	**Circulatory disease**	
ICD-9	459	Other disorders of circulatory system
		459.0-459.9
ICD-10	I00-I02	Acute rheumatic fever
	I05-I09	Chronic rheumatic heart diseases
	I10-I15	Hypertensive diseases
	I20-I25	Ischaemic heart diseases
	I26-I28	Pulmonary heart disease and diseases of pulmonary circulation
	I30-I52	Other forms of heart disease
	I60-I69	Cerebrovascular diseases
	I70-I79	Diseases of arteries, arterioles and capillaries
	I80-I89	Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified
244.4	**Hypothyroidism NOS**	
ICD-9	244.9	Unspecified hypothyroidism

ICD-10	E03.9	Hypothyroidism, unspecified
246	**Disorders of thyroid**	
ICD-9	246	Disorders of thyrocalcitonin secretion
		246.0-246.9
ICD-10	E00	Congenital iodine-deficiency syndrome
		E00.0-E00.9
	E01	Iodine-deficiency-related thyroid disorders and allied conditions
		E01.1-E01.8
	E02	Subclinical iodine-deficiency hypothyroidism
	E03	Other hypothyroidism
		E03.0-E03.9
	E04	Other nontoxic goitre
		E04.0-E04.9
	E05	Thyrotoxicosis [hyperthyroidism]
		E05.0-E05.9
	E06	Thyroiditis
		E06.0-E06.9
	E07	Other disorders of thyroid
		E07.0-E07.9
557.1	**Coeliac disease**	
ICD-9	579	Coeliac disease
	579.1	Tropical sprue
ICD-10	K90.0	Coeliac disease
	K90.1	Tropical sprue
275.1	**Disorders of iron metabolism**	
ICD-9	275	Disorders of iron metabolism
ICD-10	E83.1	Disorders of iron metabolism
272.11	**Hypercholesterolaemia**	
ICD-9	272	Pure hypercholesterolaemia
ICD-10	E78.0	Pure hypercholesterolaemia
277	**Disorders of metabolism**	
ICD-9	277	Other and unspecified disorders of metabolism
		277.0-277.9
	783.9	Other symptoms concerning nutrition, metabolism, and development
	794.7	Basal metabolism
ICD-10	C96.0	Multifocal and multisystemic (disseminated) Langerhans-cell histiocytosis
	E43	Unspecified severe protein-energy malnutrition
	E70	Disorders of aromatic amino-acid metabolism
		E70.1-E70.9
	E71	Disorders of branched-chain amino-acid metabolism and fatty-acid metabolism
	E72	Other disorders of amino-acid metabolism

	E73	Lactose intolerance
	E74	Other disorders of carbohydrate metabolism
	E75	Disorders of sphingolipid metabolism and other lipid storage disorders
	E76	Disorders of glycosaminoglycan metabolism
	E77	Disorders of glycoprotein metabolism
	E78	Disorders of lipoprotein metabolism and other lipidaemias
	E79	Disorders of purine and pyrimidine metabolism
	E80	Disorders of porphyrin and bilirubin metabolism
	E83	Disorders of mineral metabolism
	E84	Cystic fibrosis
	E85	Amyloidosis
	E86	Volume depletion
	E87	Other disorders of fluid, electrolyte and acid-base balance
	E88	Other metabolic disorders
	E89	Postprocedural endocrine and metabolic disorders, not elsewhere classified
	E90	Nutritional and metabolic disorders in diseases classified elsewhere
366	**Cataract**	
ICD-9	366	Cataract
		366.0-366.9
	998.82	Cataract fragments in eye following cataract surgery
ICD-10	V43.1	Lens
	V45.61	Cataract extraction status
	H26	Other cataract
		H26.0-H26.9
	H28	Cataract and other disorders of lens in diseases classified elsewhere
		H28.0-H28.8
	Z96.1	Presence of intraocular lens
960	**Poisoning by antibiotics**	
ICD-9	960	Poisoning by antibiotics
		960.1-960.9
	961.8	Other anti-mycobacterial drugs
	V14.0	Personal history of allergy to Penicillin
	V14.1	Personal history of allergy to other antibiotic agent
	E856	Accidental poisoning by antibiotics
	E930	Adverse effects in therapeutic use of antibiotics
		E930.0- E930.9
	E931.8	Adverse effects in therapeutic use of anti-mycobacterial drugs
	E933.4	Adverse effects in therapeutic use of Penicillinase
ICD-10	T36	Poisoning by systemic antibiotics
		T36.5- T36.9
	T37.1	Poisoning by anti-mycobacterial drugs

	T37.8	Poisoning by other specified systemic anti-infectives and antiparasitics
	Z88.1	Personal history of allergy to other antibiotic agents
471	**Nasal polyps**	
ICD-9	471	Nasal polyps
		471.0-471.9
ICD-10	J33	Nasal polyp
		J33.0-J33.9
512	**Other diseases of respiratory system**	
ICD-9	519	Other diseases of respiratory system
		519.0-519.9
	786	Symptoms involving respiratory system and other chest symptoms
		786.0-786.9
ICD-10	J95	Postprocedural respiratory disorders, not elsewhere classified
		J95.0-J95.9
	J96	Respiratory failure, not elsewhere classified
		J96.0-J96.9
	J98	Other respiratory disorders
		J98.0-J98.9
	J99	Respiratory disorders in diseases classified elsewhere
		J99.0-J99.9
535	**Gastritis and duodenitis**	
ICD-9	535	Gastritis and duodenitis
		535.0-535.7
ICD-10	K29	Gastritis and duodenitis
		K29.0-K29.9
211	**Benign neoplasm of digestive system**	
ICD-9	211	Benign neoplasm of other parts of digestive system
		211.0-211.9
ICD-10	D13	Benign neoplasm of other and ill-defined parts of digestive system
		D13.0-D13.9
	D19.1	Mesothelial tissue of peritoneum
	K31.7	Polyp of stomach and duodenum
669	**Complications of labour and delivery**	
ICD-9	667	Retained placenta without haemorrhage
	669	Other complications of labour and delivery, not elsewhere classified
		669.0-669.9
	674	Other and unspecified complications of the puerperium, not elsewhere classified
		674.0-674.9
	677	Late effect of complication of pregnancy, childbirth, and the puerperium
ICD-10	O26.5	Maternal hypotension syndrome

	O60	Preterm labour and delivery
		O60.0- O60.9
	O61	Failed induction of labour
		O61.0- O61.9
	O62	Abnormalities of forces of labour
		O62.0- O62.9
	O63	Long labour
		O63.0- O63.9
	O64	Obstructed labour due to malposition and malpresentation of foetus
		O64.0- O64.9
	O65	Obstructed labour due to maternal pelvic abnormality
		O65.0- O65.9
	O66	Other obstructed labour
		O66.0- O66.9
	O67	Labour and delivery complicated by intrapartum haemorrhage, not elsewhere classified
		O67.0- O67.9
	O68	Labour and delivery complicated by foetal stress
		O68.0- O68.9
	O69	Labour and delivery complicated by umbilical cord complications
		O69.0- O69.9
	O70	Perineal laceration during delivery
		O70.0- O70.9
	O71	Other obstetric trauma
		O71.0- O71.9
	O72	Postpartum haemorrhage
		O72.0- O72.9
	O73	Retained placenta and membranes, without haemorrhage
		O73.0- O73.9
	O74	Complications of anaesthesia during labour and delivery
		O74.0- O74.9
	O75	Other complications of labour and delivery, not elsewhere classified
		O75.0- O75.9
	O81	Single delivery by forceps and vacuum extractor
		O81.3-O81.5
	O82	Single delivery by caesarean section
	O83	Other assisted single delivery
	O90.4	Postpartum acute renal failure
689	**Disorder of skin and subcutaneous tissue**	
ICD-9	709	Other disorders of skin and subcutaneous tissue
ICD-10	L00-L08	Infections of the skin and subcutaneous tissue

	L10-L14	Bullous disorders
	L20-L30	Dermatitis and eczema
	L40-L45	Papulosquamous disorders
	L50-L54	Urticarial and erythema
	L55-L59	Radiation-related disorders of the skin and subcutaneous tissue
	L60-L75	Disorders of skin appendages
	L80-L99	Other disorders of the skin and subcutaneous tissue

Supplementary Table 5 - 7: Sex-stratified MR IVW analysis and MR Egger analysis.

Outcomes	Male-specific effect									Female-specific effect								
	MR IVW				MR Egger					MR IVW				MR Egger				
	Beta	**SE**	**OR (95%CI)**	**P *effect***	**Beta**	**SE**	**OR (95%CI)**	**P *effect***	**P *pleiotropy***	**Beta**	**SE**	**OR (95%CI)**	**P *effect***	**Beta**	**SE**	**OR (95%CI)**	**P *effect***	**P *pleiotropy***
Gout	1.77	0.12	5.85 (4.59, 7.46)	2.22E-15	2.06	0.28	7.81 (4.39, 13.90)	4.87E-08	0.20	1.23	0.30	3.42 (1.84, 6.38)	3.40E-04	1.10	0.47	3.01 (1.16, 7.81)	0.03	0.68
Inflammatory polyarthropathies	0.57	0.07	1.76 (1.53, 2.04)	5.41E-09	0.64	0.13	1.90 (1.45, 2.48)	3.76E-05	0.49	-0.04	0.06	0.96 (0.85, 1.07)	0.43	-0.08	0.07	0.92 (0.79, 1.07)	0.28	0.46
Essential hypertension	0.09	0.04	1.10 (1.02, 1.19)	0.02	-0.07	0.08	0.93 (0.78, 1.10)	0.37	0.02	0.07	0.03	1.08 (1.00, 1.15)	0.04	-0.06	0.07	0.94 (0.82, 1.08)	0.39	0.01
Hypertensive disease	0.09	0.04	1.10 (1.02, 1.19)	0.02	-0.07	0.08	0.93 (0.78, 1.10)	0.37	0.02	0.07	0.03	1.08 (1.00, 1.16)	0.04	-0.06	0.07	0.94 (0.82, 1.08)	0.40	0.00
Myocardial infarction	0.15	0.07	1.17 (1.01, 1.34)	0.03	0.03	0.15	1.03 (0.76, 1.39)	0.84	0.29	0.18	0.11	1.20 (0.96, 1.50)	0.10	0.08	0.15	1.08 (0.79, 1.47)	0.62	0.28
Coeliac disease	0.35	0.26	1.42 (0.83, 2.43)	0.19	0.03	0.60	1.03 (0.31, 3.49)	0.96	0.51	0.27	0.16	1.31 (0.94, 1.84)	0.11	0.22	0.28	1.24 (0.70, 2.22)	0.45	0.77
Disorders of metabolism	0.10	0.04	1.10 (1.01, 1.20)	0.04	0.04	0.09	1.04 (0.86, 1.26)	0.70	0.44	0.37	0.38	1.44 (0.66, 3.14)	0.34	0.08	0.75	1.08 (0.24, 4.96)	0.92	0.55
Coronary atherosclerosis	0.05	0.05	1.05 (0.95, 1.16)	0.32	-0.03	0.10	0.97 (0.79, 1.20)	0.79	0.36	0.14	0.06	1.15 (1.02, 1.30)	0.02	0.08	0.10	1.09 (0.88, 1.34)	0.43	0.38
Ischaemic heart disease	0.04	0.05	1.04 (0.95, 1.15)	0.39	-0.03	0.10	0.97 (0.79, 1.19)	0.77	0.38	0.14	0.06	1.15 (1.02, 1.30)	0.02	0.07	0.10	1.08 (0.87, 1.33)	0.47	0.33
Angina pectoris	0.03	0.06	1.03 (0.90, 1.17)	0.66	-0.02	0.12	0.98 (0.77, 1.26)	0.89	0.64	0.12	0.08	1.13 (0.97, 1.32)	0.11	0.01	0.12	1.01 (0.80, 1.29)	0.90	0.15
Atrial fibrillation and flutter	-0.03	0.07	0.97 (0.85, 1.11)	0.67	-0.20	0.11	0.81 (0.65, 1.03)	0.08	0.06	0.08	0.08	1.09 (0.92, 1.29)	0.33	-0.02	0.12	0.98 (0.77, 1.25)	0.85	0.18
Circulatory disease	0.05	0.03	1.05 (0.98, 1.12)	0.16	0.01	0.06	1.01 (0.89, 1.13)	0.92	0.39	0.03	0.03	1.03 (0.98, 1.09)	0.23	-0.04	0.05	0.96 (0.86, 1.07)	0.49	0.05
Varicose veins of lower extremity	-0.15	0.11	0.86 (0.69, 1.08)	0.19	-0.15	0.21	0.86 (0.55, 1.33)	0.48	0.97	-0.14	0.07	0.87 (0.76, 0.99)	0.04	-0.14	0.13	0.87 (0.67, 1.13)	0.29	0.98
Disorders of iron metabolism	0.15	0.30	1.16 (0.63, 2.16)	0.62	-0.42	1.05	0.66 (0.08, 5.60)	0.69	0.50	0.37	0.38	1.44 (0.66, 3.14)	0.34	0.08	0.75	1.08 (0.24, 4.96)	0.92	0.55
Hypercholesterolaemia	0.22	0.12	1.24 (0.98, 1.57)	0.07	0.20	0.19	1.22 (0.82, 1.81)	0.32	0.91	0.06	0.05	1.06 (0.95, 1.18)	0.29	-0.04	0.12	0.96 (0.75, 1.23)	0.74	0.21
Hypothyroidism	0.20	0.13	1.23 (0.95, 1.59)	0.12	0.21	0.26	1.23 (0.73, 2.08)	0.42	0.98	0.06	0.05	1.06 (0.95, 1.18)	0.29	-0.04	0.12	0.96 (0.75, 1.23)	0.74	0.21
Disorders of thyroid	0.20	0.12	1.22 (0.96, 1.55)	0.09	0.22	0.22	1.25 (0.80, 1.97)	0.32	0.90	0.04	0.05	1.04 (0.94, 1.15)	0.48	-0.04	0.11	0.96 (0.77, 1.20)	0.71	0.28
Benign neoplasm of digestive system	0.06	0.14	1.06 (0.79, 1.41)	0.70	-0.01	0.26	0.99 (0.59, 1.68)	0.97	0.75	-0.11	0.10	0.89 (0.73, 1.10)	0.27	-0.09	0.18	0.91 (0.63, 1.33)	0.63	0.85

Outcomes	Male-specific effect									Female-specific effect								
	MR IVW				MR Egger					MR IVW				MR Egger				
	Beta	**SE**	**OR (95%CI)**	**P *effect***	**Beta**	**SE**	**OR (95%CI)**	**P *effect***	**P *pleiotropy***	**Beta**	**SE**	**OR (95%CI)**	**P *effect***	**Beta**	**SE**	**OR (95%CI)**	**P *effect***	**P *pleiotropy***
Gastritis and duodenitis	-0.04	0.07	0.96 (0.82, 1.11)	0.56	-0.03	0.15	0.97 (0.71, 1.33)	0.83	0.93	-0.03	0.06	0.97 (0.86, 1.09)	0.62	-0.07	0.08	0.93 (0.79, 1.11)	0.41	0.45
Nasal polyps	0.23	0.14	1.25 (0.95, 1.65)	0.11	0.41	0.28	1.51 (0.86, 2.68)	0.15	0.40	-0.23	0.17	0.79 (0.56, 1.11)	0.17	-0.37	0.27	0.69 (0.40, 1.19)	0.17	0.41
Cataract	-0.01	0.08	0.99 (0.84, 1.16)	0.88	-0.13	0.16	0.88 (0.63, 1.23)	0.43	0.38	-0.02	0.06	0.98 (0.86, 1.11)	0.74	-0.10	0.10	0.90 (0.74, 1.10)	0.30	0.20
Poisoning by antibiotics	-0.15	0.23	0.86 (0.53, 1.39)	0.52	0.03	0.45	1.03 (0.41, 2.56)	0.95	0.62	-0.14	0.11	0.87 (0.70, 1.09)	0.21	0.00	0.16	1.00 (0.71, 1.40)	0.99	0.21
Complications of labour and delivery	NA	NA	NA	NA	NA	NA	NA	NA	NA	-0.11	0.07	0.90 (0.78, 1.04)	0.14	-0.18	0.12	0.83 (0.65, 1.06)	0.13	0.31
Other diseases of respiratory system	-0.14	0.11	0.87 (0.69, 1.09)	0.21	-0.35	0.24	0.70 (0.43, 1.14)	0.15	0.28	0.38	0.11	1.46 (1.15, 1.84)	2.62E-03	0.51	0.15	1.66 (1.22, 2.26)	2.27E-03	0.20
Disorder of skin and subcutaneous tissue	0.03	0.05	1.03 (0.93, 1.13)	0.60	0.003	0.09	1.00 (0.84, 1.20)	0.97	0.74	0.003	0.04	1.00 (0.92, 1.09)	0.93	0.04	0.07	1.04 (0.90, 1.20)	0.60	0.46

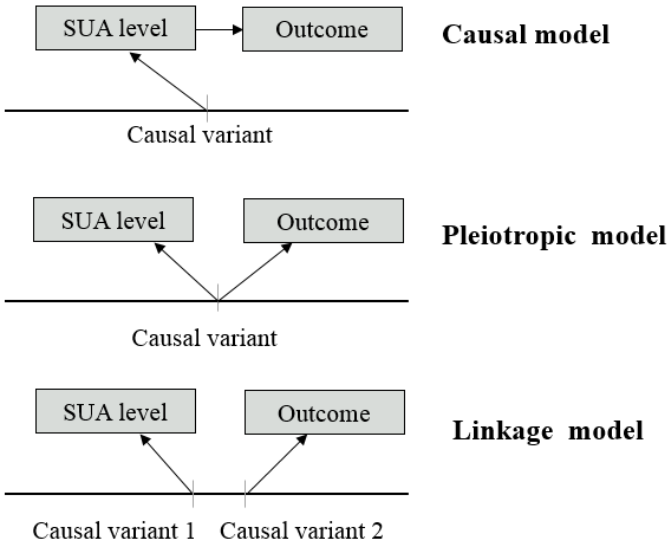
In sex-stratified MR IVW analysis of 25 disease groups/outcomes (**Supplementary Table 5-7**), 3 disease groups/outcomes (gout, hypertensive disease, and essential hypertension) showed potential causal link with SUA level in both men and women, 3 disease groups/outcomes (inflammatory polyarthropathies, myocardial infarction, and disorders of metabolism) had potential causal link in men, and 4 disease groups/outcomes (coronary atherosclerosis, ischaemic heart disease, varicose veins of lower extremity and other diseases of respiratory system) had potential causal link only in women. The sex-stratified MR Egger analysis suggested causal effect of SUA level on gout in both men (OR=7.81, 95%CI: 4.39 to 13.90, P_{effect}= 4.87×10^{-8}, $P_{pleiotropy}$=0.20) and women (OR=3.01, 95%CI: 1.16 to 7.81, P_{effect}= 0.03, $P_{pleiotropy}$=0.68), on inflammatory polyarthropathies in men (OR=1.90, 95%CI: 1.45 to 2.48, P_{effect}= 3.76×10^{-5}, $P_{pleiotropy}$=0.49), and on other diseases of respiratory system in women (OR=1.66, 95%CI: 1.22 to 2.26, P_{effect}= 2.27×10^{-3}, $P_{pleiotropy}$=0.20).

Supplementary Table 5 - 8: A sensitivity analysis of MR for gout cases defined by multiple criteria.

Case ascertainment	No. of cases*	MR IVW			MR Egger		
		OR (95%CI)	P *effect*	Power	OR (95%CI)	P *effect*	Power
Hospital-diagnosed gout	1,003	4.88 (3.91, 6.09)	3.55E-15	1.00	4.58 (2.72, 7.72)	1.76E-06	1.00
Self-reported gout	1,769	9.78 (8.22, 9.79)	0.00E+00	1.00	14.5 (10.1, 14.9)	2.89E-15	1.00
Hospital-diagnosed or self-reported gout	2,274	8.26 (7.04, 9.60)	0.00E+00	1.00	12.6 (9.06, 17.3)	6.66E-16	1.00

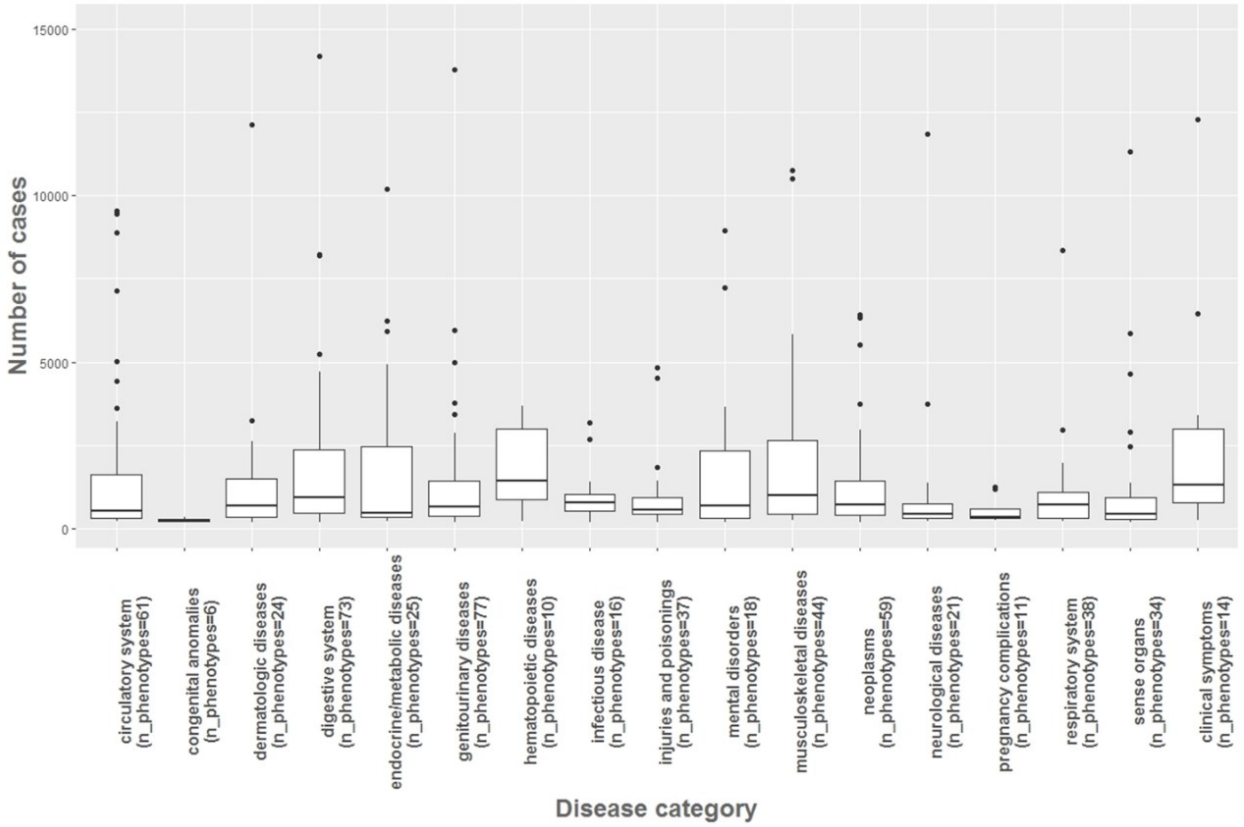
*Note the number of study population is 120,091.

Online Supplementary Figures

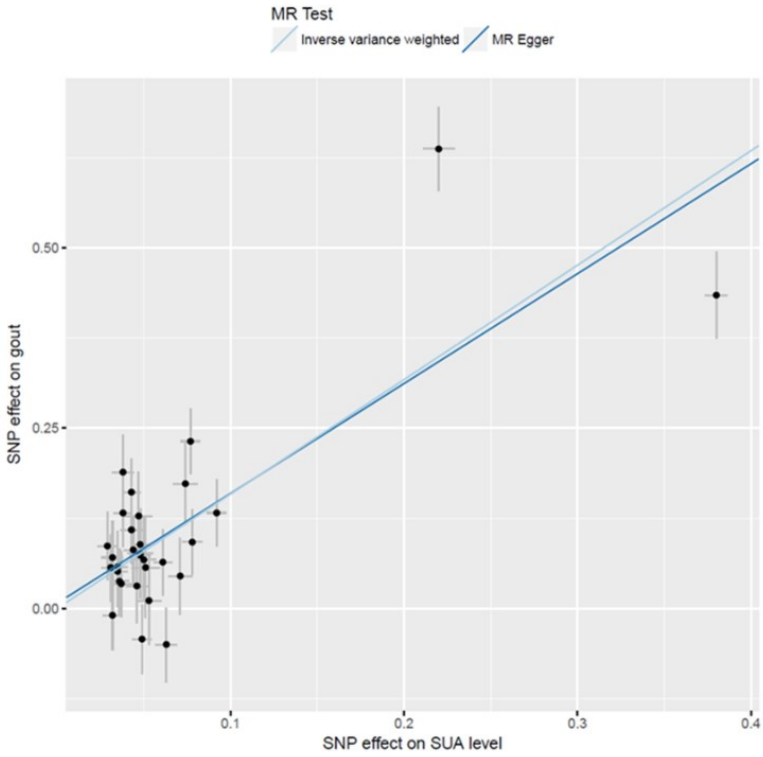


Supplementary Figure 5 - 1: Three possible explanations for the PheWAS association between SUA level and disease outcome through genotypes.

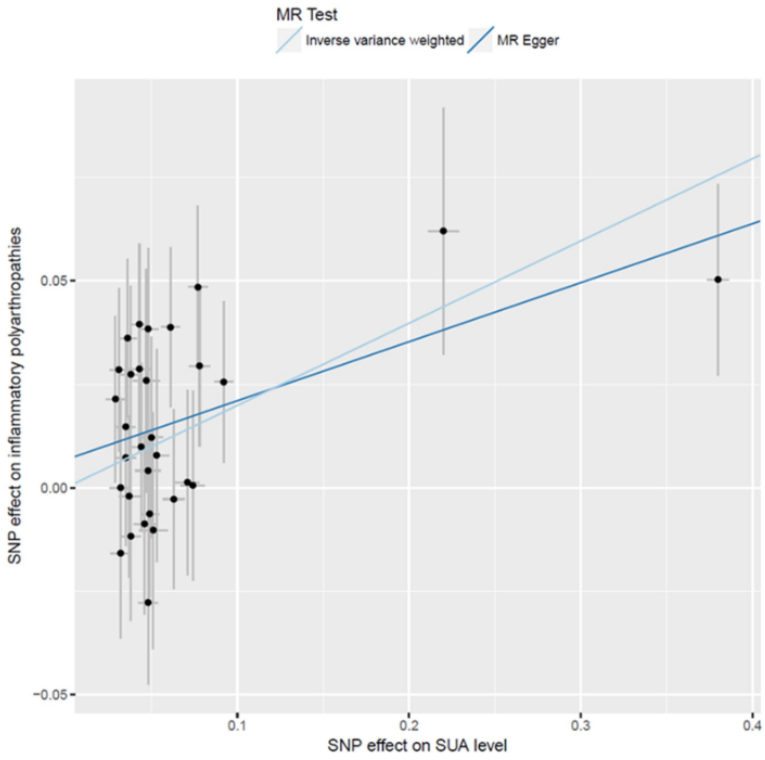
(Reshaped based on the publication of Zhu *et al*).



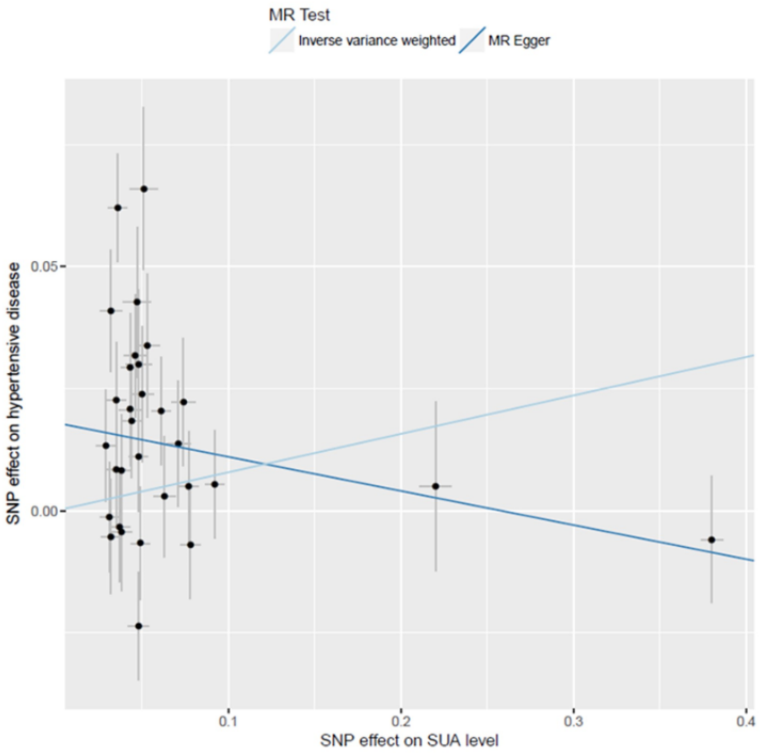
Supplementary Figure 5 - 2: The number of phenotypes and median number of cases in each disease category (outliers>15K are not plotted).



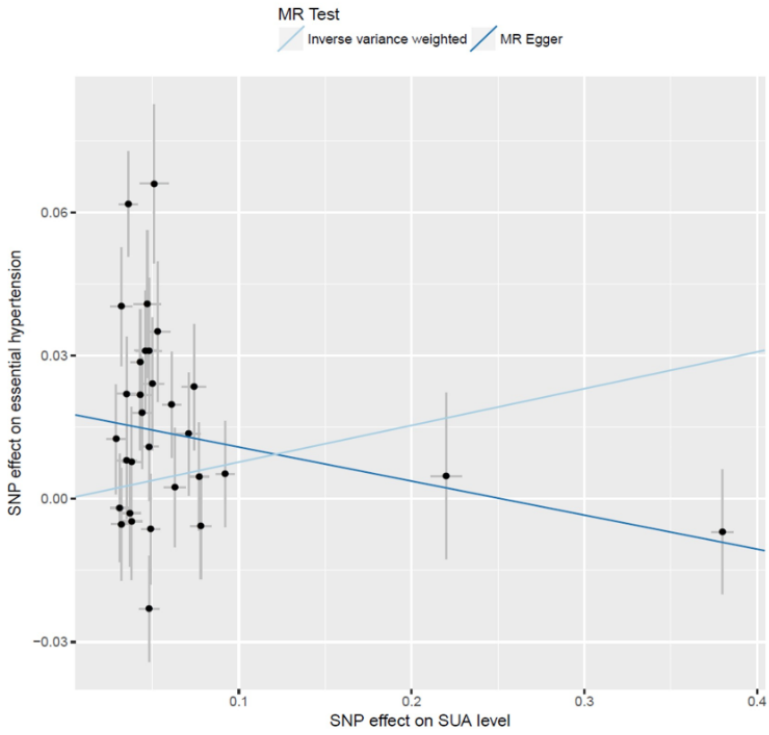
Supplementary Figure 5 - 3: A scatter plot of the SNP effect on SUA level against the SNP effect on gout.



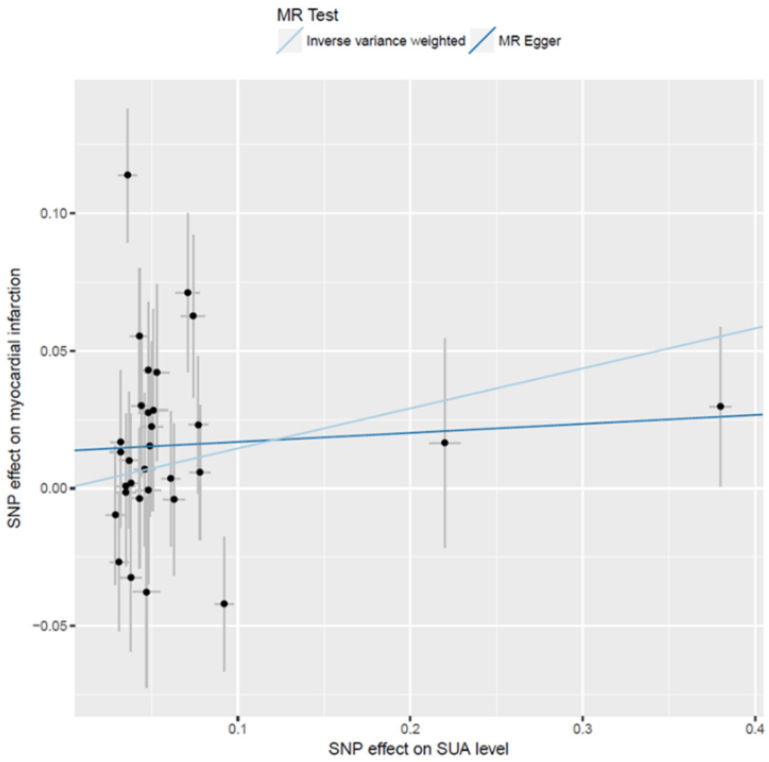
Supplementary Figure 5 - 4: A scatter plot of the SNP effect on SUA level against the SNP effect on inflammatory polyarthropathies.



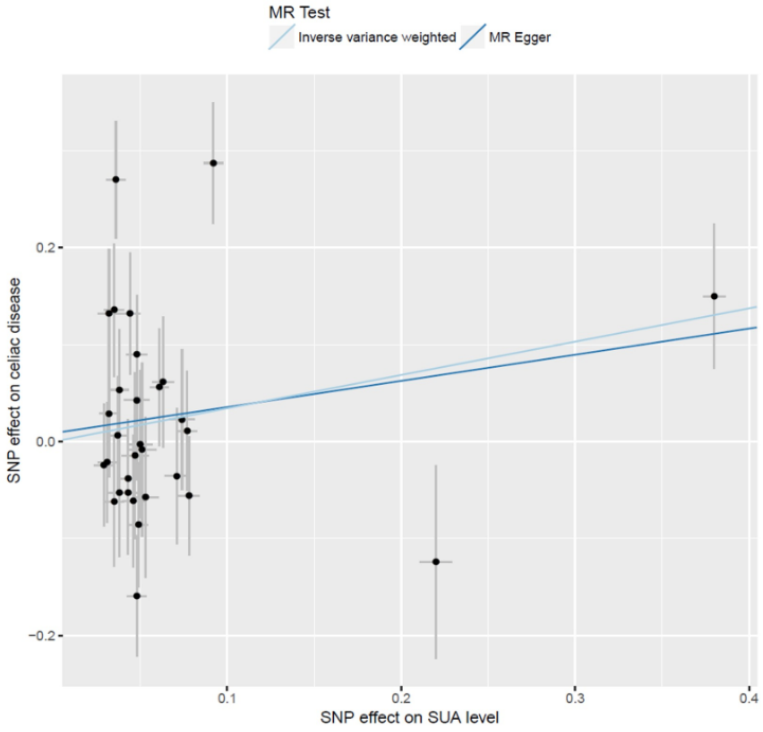
Supplementary Figure 5 - 5: A scatter plot of the SNP effect on SUA level against the SNP effect on hypertensive disease.



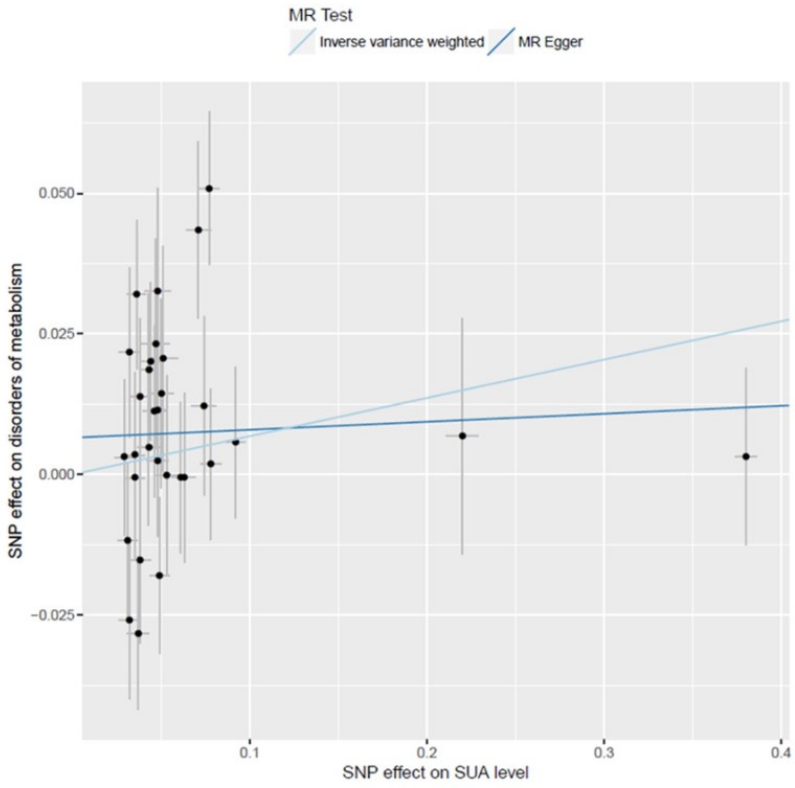
Supplementary Figure 5 - 6: A scatter plot of the SNP effect on SUA level against the SNP effect on essential hypertension.



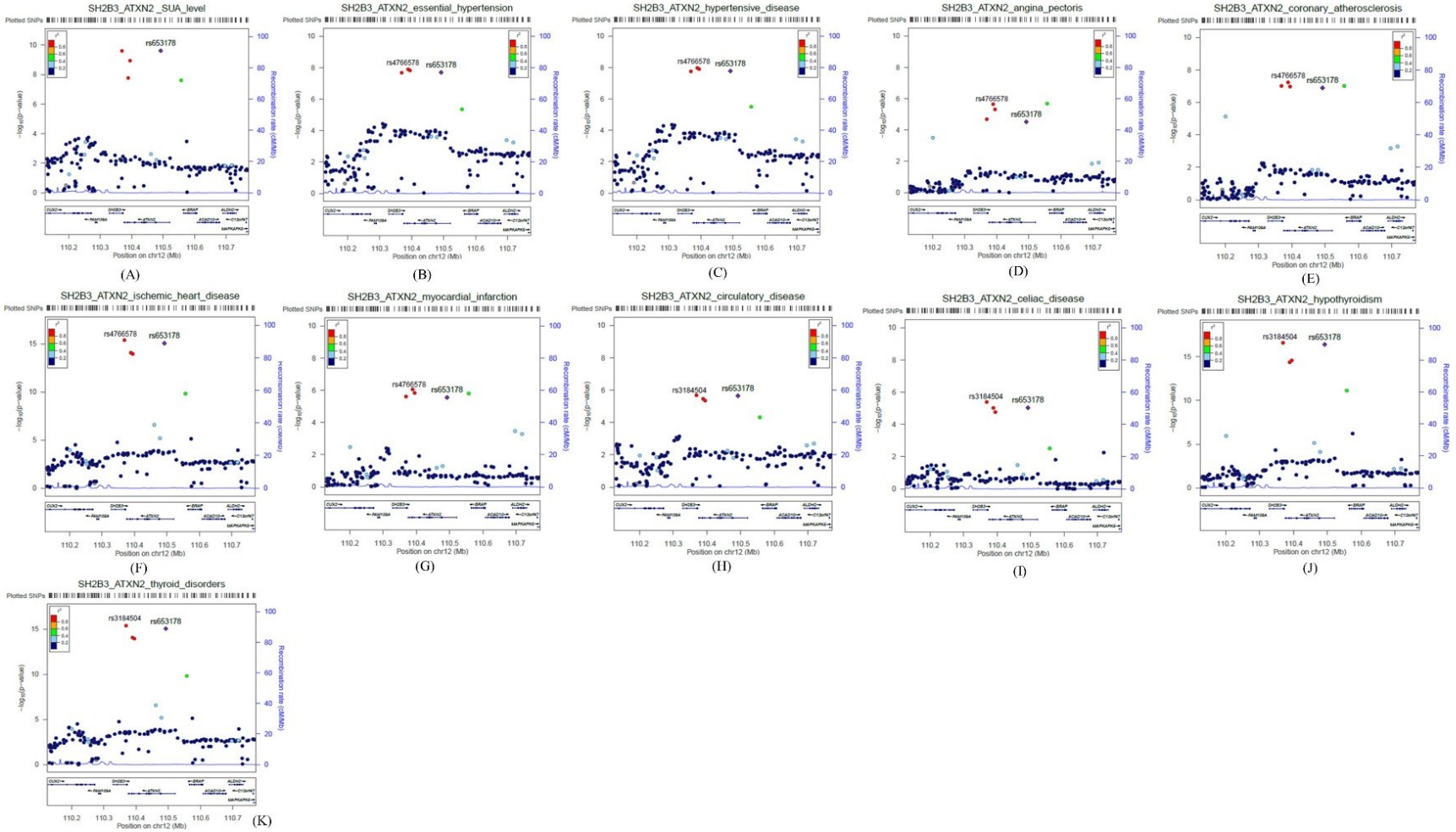
Supplementary Figure 5 - 7: A scatter plot of the SNP effect on SUA level against the SNP effect on myocardial infarction.



Supplementary Figure 5 - 8: A scatter plot of the SNP effect on SUA level against the SNP effect on coeliac disease.

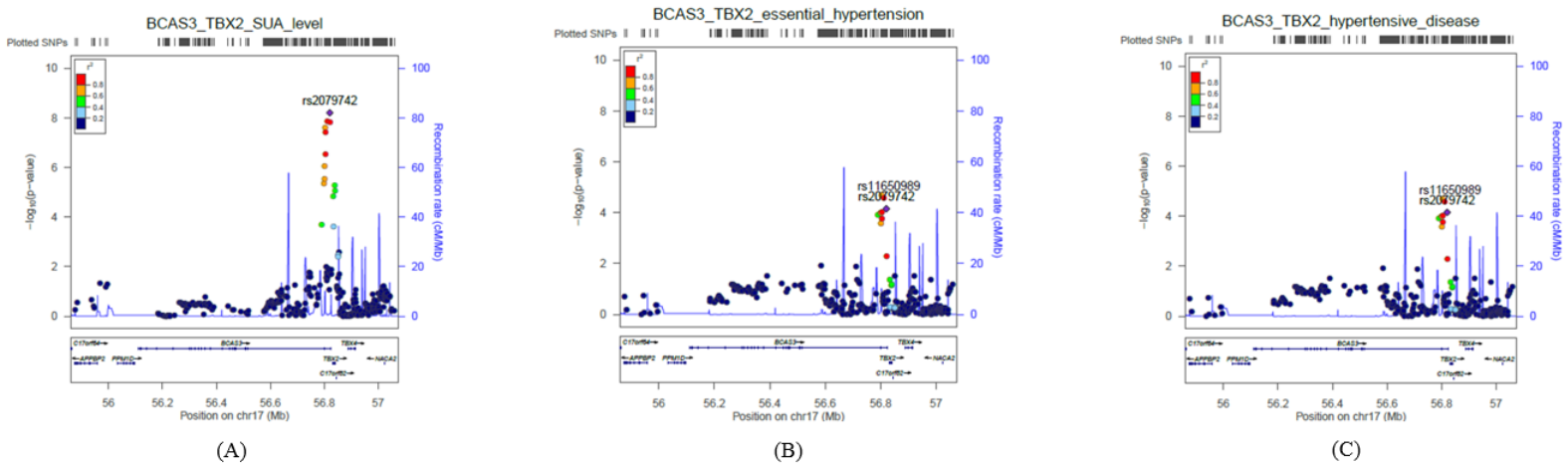


Supplementary Figure 5 - 9: A scatter plot of the SNP effect on SUA level against the SNP effect on disorders of metabolism.



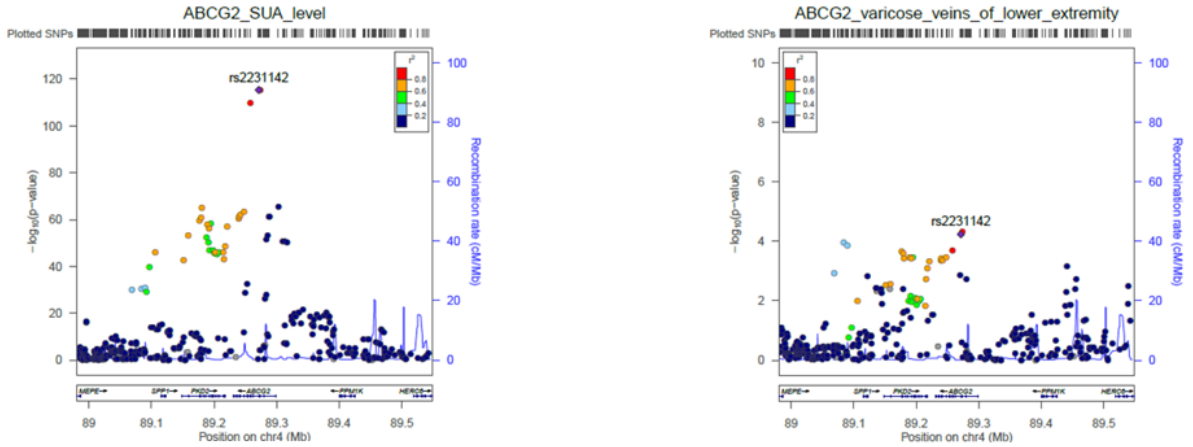
Supplementary Figure 5 - 10: Regional association plots of *SH2B3_ATXN2* locus with associated phenotypes.

Results of regional association were plotted as –log10 P-value for all SNPs located within a±250 kb window around *SH2B3_ATXN2* ordered by the chromosome position. The LD between the index SNP (rs653178) and other SNPs in this region were indicated by using color scheme. Location of genes were given by blue narrows at the bottom, pointing the direction of transcription. Both the index SNP (rs653178) and the most significant SNPs associated with the phenotypes (top SNP with the smallest P-value) were tagged. Results were shown for (A) the regional association of the *SH2B3_ATXN2* locus with SUA level; (B) the regional association of the *SH2B3_ATXN2* locus with essential hypertension (top SNP: rs4766578; p value of HEIDI test = 0.99); (C) the regional association of the *SH2B3_ATXN2* locus with hypertensive disease (top SNP: rs4766578; p value of HEIDI test = 0.99); (D) the regional association of the *SH2B3_ATXN2* locus with angina pectoris (top SNP: rs4766578; p value of HEIDI test = 0.20); (E) the regional association of the *SH2B3_ATXN2* locus with coronary atherosclerosis (top SNP: rs4766578; p value of HEIDI test = 0.91); (F) the regional association of the *SH2B3_ATXN2* locus with ischaemic heart disease (top SNP: rs4766578; p value of HEIDI test = 0.93); (G) the regional association of the *SH2B3_ATXN2* locus with myocardial infarction (top SNP: rs4766578; p value of HEIDI test = 0.06); (H) the regional association of the *SH2B3_ATXN2* locus with circulatory disease (top SNP: rs3184504; p value of HEIDI test = 0.99); (I) the regional association of the *SH2B3_ATXN2* locus with coeliac disease (top SNP: rs3184504; p value of HEIDI test = 0.67); (J) the regional association of the *SH2B3_ATXN2* locus with hypothyroidism (top SNP: rs3184504; p value of HEIDI test = 0.57); (K) the regional association of the *SH2B3_ATXN2* locus with thyroid disorders (top SNP: rs3184504; p value of HEIDI test = 0.88).



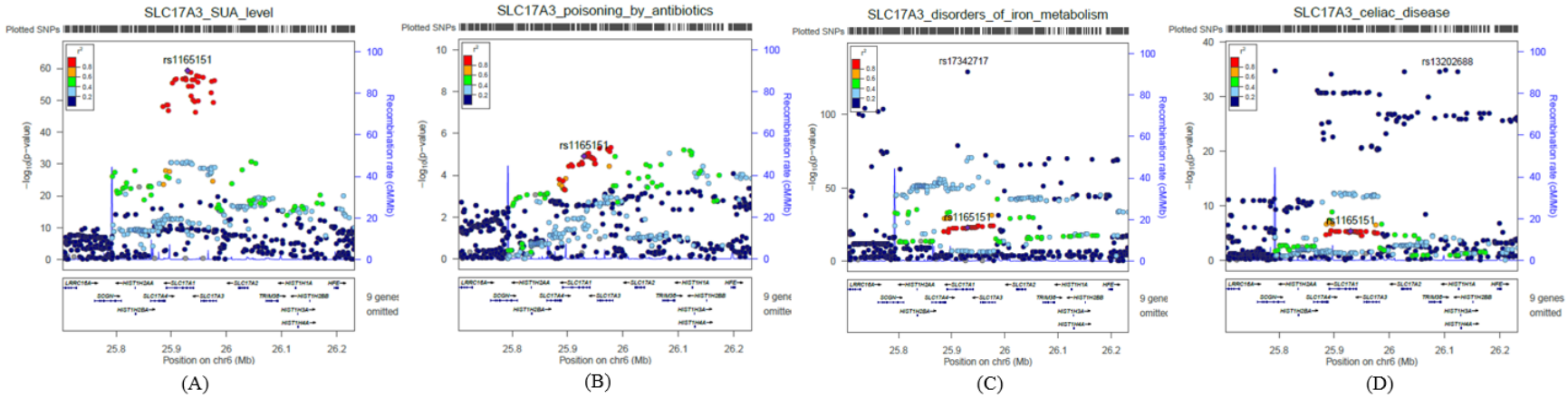
Supplementary Figure 5 - 11: Regional association plots of the *BCAS3_TBX2* locus with associated phenotypes.

Results of regional association were plotted as –log10 P-value for all SNPs located within a ±250 kb window around *BCAS3_TBX2* ordered by the chromosome position. The LD between the index SNP (rs2079742) and other SNPs in this region were indicated with the color scheme. Location of genes were given by blue narrows at the bottom, pointing the direction of transcription. Both the index SNP (rs2079742) and the most significant SNP (with the smallest P-value) associated with the phenotypes (rs11650989). Results were shown for (A) the regional association of the *BCAS3_TBX2* locus with SUA level; (B) the regional association of the *BCAS3_TBX2* locus with essential hypertension (top SNP: rs11650989; p value of HEIDI test = 0.09); (C) the regional association of the *BCAS3_TBX2* locus with hypertensive disease (top SNP: rs11650989; p value of HEIDI test = 0.10).



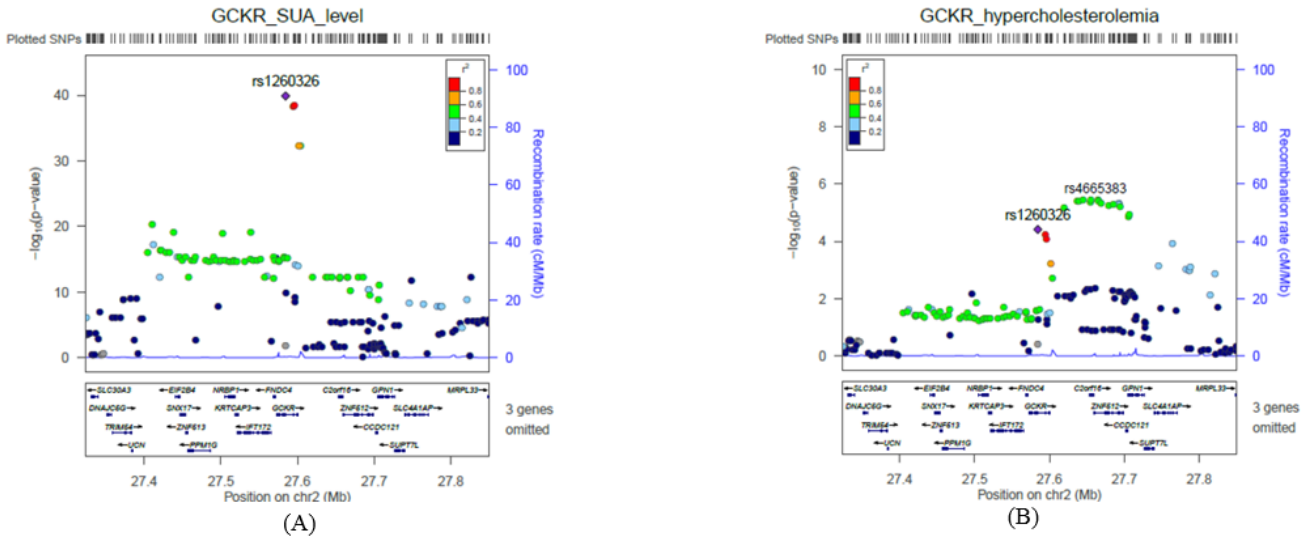
Supplementary Figure 5 - 12: Regional association plots of the *ABCG2* locus with associated phenotypes.

Results of regional association were plotted as –log10 P-value for all SNPs located within a ±250 kb window around *ABCG2* ordered by the chromosome position. The LD between the index SNP (rs2231142) and other SNPs in this region were indicated with the color scheme. Location of genes were given by blue narrows at the bottom, pointing the direction of transcription. Both the index SNP (rs2231142) and the most significant SNP (with the smallest P-value) associated with the phenotypes were tagged. Results were shown for (A) the regional association of *ABCG2* locus with SUA level; (B) the regional association of the *ABCG2* locus with varicose veins of lower extremity (top SNP: rs2231142; p value of HEIDI test = 0.32).



Supplementary Figure 5 - 13: Regional association plots of the *SLC17A3* locus with associated phenotypes.

Results of regional association were plotted as –log10 P-value for all SNPs located within a ±250 kb window around *SLC17A3* ordered by the chromosome position. The LD between the index SNP (rs1165151) and other SNPs in this region were indicated with the color scheme. Location of genes were given by blue narrows at the bottom, pointing the direction of transcription. Both the index SNP (rs1165151) and the most significant SNP (with the smallest P-value) associated with the phenotypes were tagged. Results were shown for (A) the regional association of the *SLC17A3* locus with SUA level; (B) the regional association of the *SLC17A3* locus with poisoning by antibiotics (top SNP: rs1165205; p value of HEIDI test = 0.26); (C) the regional association of *SLC17A3* locus with disorders of iron metabolism (top SNP: rs17342717; p value of HEIDI test = 5.54×10^{-28}); (D) the regional association of the *SLC17A3* locus with coeliac disease (top SNP: rs13202688; p value of HEIDI test = 6.51×10^{-16}).



Supplementary Figure 5 - 14: Regional association plots of the *GCKR* locus with associated phenotypes.

Results of regional association were plotted as –log10 P-value for all SNPs located within a ±250 kb window around *GCKR* ordered by the chromosome position. The LD between the index SNP (rs1260326) and other SNPs in this region were indicated with the color scheme. Location of genes were given by blue narrows at the bottom, pointing the direction of transcription. Both the index SNP (rs1260326) and the most significant SNP (with the smallest P-value) associated with the phenotypes (rs4665383) were tagged. Results were shown for (A) the regional association of *GCKR* locus with SUA level; (B) the regional association of the *GCKR* locus with hypercholesterolaemia (top SNP: rs4665383; p value of HEIDI test = 3.27×10^{-11}).

6 PWMR ANALYSIS: FULL UK BIOBANK DATA

6.1 Summary

This chapter presents a Phenome-wide Mendelian randomisation (PWMR) study by using data from an unrelated White British subset (n=339,256) selected from the full UK Biobank cohort (the selection process is presented in *Chapter 4, Section 4.3.1 "Study population selection"*). The present study aims to extend the prior findings by combining genetic risk loci of urate into a weighted GRS, investigating more disease outcomes, assessing their associations with more cases (>3-fold), examining consistency of findings across two different phenotyping models, and replicating the findings by performing two-sample MR in different populations.

A weighted polygenic risk score (GRS), incorporating effect estimates of multiple genetic risk loci, was employed as a proxy of serum urate level. The framework of phenome was defined by using both the PheCODE schema (also used in the previous MR-PheWAS) and a novel Bayesian analysis framework, termed TreeWAS: tree-structured phenotypic model. To validate the findings, identified associations were further examined in the MR-base database for replication in different populations. Sensitivity analysis examining the pleiotropic effects of urate genetic risk loci on a set of metabolic traits was performed to explore any causal effect and pleiotropic association.

The PheWAS analysis based on the PheCODE schema examined the association between a weighted GRS of SUA level and 1,431 disease outcomes and identified 13 phecodes that had p value less than the significance threshold of PheWAS ($p<3.35\times10^{-4}$). These phecodes represent 4 disease groups: inflammatory polyarthropathies (OR=1.28; 95% CI: 1.21 to 1.35; $p=4.97\times10^{-19}$), hypertensive disease (OR=1.08; 95% CI: 1.05 to 1.11; $p=6.02\times10^{-7}$), circulatory disease (OR=1.05; 95% CI: 1.02 to 1.07; $p=3.29\times10^{-4}$) and metabolic disorders (OR=1.07; 95% CI: 1.03 to 1.11; $p= 3.33\times10^{-4}$), and 9 disease outcomes: gout (OR=5.37; 95% CI: 4.67 to 6.18; $p= 4.27\times10^{-123}$), gouty arthropathy (OR=5.11; 95% CI: 2.45 to 10.66; $p=1.39\times10^{-5}$), pyogenic arthritis (OR=2.10; 95% CI: 1.41 to 3.14; $p=2.87\times10^{-4}$), essential hypertension (OR=1.08; 95% CI: 1.05 to 1.11; $p=6.62\times10^{-7}$), coronary atherosclerosis (OR=1.10; 95% CI: 1.05 to 1.15; $p=1.17\times10^{-5}$), ischaemic heart disease (OR=1.10, 95% CI: 1.05 to 1.15; $p=1.73\times10^{-5}$), chronic ischaemic heart disease (OR=1.10, 95% CI: 1.05 to 1.15; $p=1.52\times10^{-5}$), myocardial infarction (OR=1.15, 95% CI=1.07 to 1.23, $p=5.23\times10^{-5}$), and hypercholesterolaemia (OR=1.08, 95% CI: 1.04 to 1.13, $p=3.34\times10^{-4}$). In the Bayesian analysis framework, containing 10,750 diagnostic terms, a total of 27 parent/child nodes of

ICD-10 terms were identified with a PP (posterior probability) ⩾0.95. They were clustered mainly in five branches of the hierarchical tree structure: (i) block M10 (gout, OR=5.16, 95%CI: 4.55 to 5.84; PP=1.00); (ii) block I10-I15 (hypertensive disease, OR=1.07, 95%CI: 1.06 to 1.08; PP>0.99); (iii) block I20-I25 (ischaemic heart diseases, OR=1.07, 95%CI: 1.06 to 1.08; PP>0.99); (iv) block I30-I52 (other forms of heart disease, OR=1.07, 95%CI: 1.06 to 1.08; PP>0.99); (v) block I60-I69 (cerebrovascular diseases, OR=1.07, 95%CI: 1.06 to 1.08; PP>0.99). Findings from PheWAS and TreeWAS were generally consistent in their associations with gout, hypertensive disease, and heart diseases, while 14 more sub-phenotypes were identified from TreeWAS.

MR IVW analysis successfully replicated the association between urate and the risk of gout, CHD, myocardial infarction and decreased level of HDL-c in different populations by analysing various GWAS consortia summary data that are included in the MR-base database. However, the MR Egger analysis indicated the existence of unbalanced genetic pleiotropy on the observed associations between urate and cardiovascular/metabolic diseases. When balancing out the potential pleiotropic effects in Egger MR, causal effect was verified for gout (OR=4.17, 95%CI: 3.03 to 5.74, $P_{effect} = 1.27 \times 10^{-9}$; $P_{pleiotropy} = 0.485$).

To further investigate the influence of pleiotropy, we re-calculated the PheWAS estimates by using a number of GRSs created based on their association with a set of metabolic traits. The GRS of urate-specific loci was only associated with gout and its upper disease group of inflammatory polyarthropathies, but not with any cardiovascular/metabolic diseases. In contrast, the GRSs of pleiotropic loci on BMI, BP, lipids and glucose showed association with both gout and the cardiovascular/metabolic diseases. When removing any group of pleiotropic loci from the creation of GRS, their association with hypertensive diseases, heart diseases, and metabolic disorders were not statistically significant.

Overall, when taken together the findings from PheWAS/TreeWAS, MR replication and sensitivity analysis, I conclude that there are robust associations between urate and a group of diseases, including gout, hypertensive diseases, heart diseases and metabolic disorders of lipids, but the causal role of urate only exists in gout. Findings in this chapter indicate that the association between urate and cardiovascular/metabolic diseases is probably due to the pleiotropic effects of genetic variants on urate and metabolic traits. These findings suggest that urate could be a good predictor for the cardiovascular/metabolic disease risk. Further investigation on therapies targeting on the shared biological pathways between urate and metabolic traits would be beneficial for both the treatment of gout and the primary prevention of cardiovascular/metabolic diseases.

For the study presented in this chapter I conducted all aspects of the research work, including designing the study, analysing the data, interpreting the findings and writing the manuscript. Specific to the contribution of co-authors, *Theodoratou, E.* and *Campbell, H.* conceived the study. *Meng, X., Wei, Q., Gifford, A., Denny, J.C.,* and *Varley, T.,* contributed to create the mapping of ICD-10/9 codes to phecode. *Spiliopoulou, A., and McKeigue, P.,* contributed important intellectual content to interpret the findings. All authors critically reviewed the manuscript and contributed important intellectual content.

A Phenome-wide Mendelian randomisation study on genetically determined serum urate levels in UK Biobank cohort

AUTHORS Xue Li[1], *MSc*; Xiangrui Meng[1], *MSc*; Athina Spiliopoulou[2], *PhD*; Maria Timofeeva[3], *PhD*; Yazhou He[1,5], *MSc*; Wei-Qi Wei[4], *MD*, *PhD*; Aliya Gifford[4], *PhD*; Tian Yang[1], *MSc*; Tim Varley[6], *BSc*; Peter Joshi[1], *PhD*; Joshua C. Denny[4], *MD*, *MS*; Paul Mckeigue[2], *PhD*; Harry Campbell[1], *PhD*; Evropi Theodoratou[1,7#], *PhD*

1 Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

2 Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

3 Colon Cancer Genetics Group, Medical Research Council Human Genetics Unit, Medical Research Council, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

4 Department of Biomedical Informatics, Vanderbilt University Medical Centre, Nashville, TN, USA

5 West China School of Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan Province, P.R. China

6 Public Health and Intelligence, NHS National Services Scotland, Edinburgh, UK

7 Edinburgh Cancer Research Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

[#] Corresponding author

Evropi Theodoratou, Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom, EH8 9AG;

E-mail: E.Theodoratou@ed.ac.uk, Tel: (+44) 131-650-6194

ABSTRACT

Objective: To explore the causal and pleiotropic associations between serum urate levels and a phenome-wide spectrum of disease outcomes using data of 339,256 individuals from the UK Biobank cohort.

Methods: A weighted polygenic risk score (GRS) of 31 genetic variants was employed as genetic proxy of serum urate levels. The phenome framework was defined by using both the PheCODE schema (termed PheWAS) and a tree-structured phenotypic model (termed TreeWAS). Significant associations from these analyses were taken forward for replication in different populations by analysing data from various GWAS consortia documented in the MR-base database. Sensitivity analyses examining the pleiotropic effects of urate genetic risk loci on a set of metabolic traits were performed to explore any causal effects and pleiotropic associations.

Results: The PheWAS analysis, examining the association with 1,431 disease outcomes, identified 13 distinct phecodes representing 4 disease groups (inflammatory polyarthropathies, hypertensive disease, circulatory disease, metabolic disorders) and 9 disease outcomes (gout, gouty arthropathy, pyogenic arthritis, essential hypertension, coronary atherosclerosis, ischaemic heart disease, chronic ischaemic heart disease, myocardial infarction, and hypercholesterolaemia) that were associated with the genetically determined serum urate levels after multiple testing correction ($p<3.35\times10^{-4}$). The Bayesian analysis of TreeWAS, examining 10,750 ICD-10 diagnostic terms, identified 27 parent/child nodes of ICD-10 terms reporting a posterior probability (PP) ≥0.95, with a few more sub-phenotypes being identified than in the PheWAS analysis. MR IVW analysis successfully replicated the associations with gout, CHD, myocardial infarction and decreased level of HDL-c, but MR Egger analysis indicated the existence of pleiotropy for most of the associations. After balancing out pleiotropic effects, a causal role of urate was verified for gout (OR=4.17, 95%CI: 3.03 to 5.74). Sensitivity analyses on the GRSs of different groups of pleiotropic loci support an inference that pleiotropic effects of genetic variants on urate and metabolic traits contribute to the observed associations with cardiovascular/metabolic diseases.

Conclusion: We conclude that there are robust associations between urate and a group of diseases, including gout, hypertensive diseases, heart diseases and metabolic disorders, but the causal role of urate is only supported in gout. Our study indicates that the observed associations between urate and cardiovascular/metabolic diseases are probably derived from

the pleiotropic effects of genetic variants on urate and metabolic traits. Further investigation of therapies targeting the shared biological pathways between urate and metabolic traits may be beneficial for the treatment of gout and the primary prevention of cardiovascular/metabolic diseases.

6.2 Introduction

The role of urate has been explored in a large number of observational studies in relation to a multitude of health outcomes (232). Apart from gout, compelling evidence supports the association between high serum urate level and the increased risk of non-crystal deposition disorders, including hypertension, cardiovascular diseases (CVDs), and metabolic syndrome (225, 226, 232, 486). Although considerable research efforts have been made in trying to understand the pathogenic role of uric acid in these non-crystal deposition disorders, its causal role has not been established (232). Therefore, it has been argued that either these associations are confounded by other risk factors, such as obesity, or they represent reverse causality (227, 522).

As in other complex traits, genetic determinants play a crucial role in the regulation of serum urate levels. Genetic studies among twins and families have reported a substantial heritable component of serum urate level with an estimated heritability of 40-70% (228, 229). The genetic determinants of serum urate level have been explored in several genome-wide association studies (GWAS) (150, 151, 230, 523) and the wealth of GWAS findings allows the application of genetic variants as instruments to help separate causal from non-causal associations, given that genotypes are generally independent of environmental exposures and the transmission of genetic information is usually unidirectional. Investigating the associations between urate genetic risk loci and disease outcomes might help provide causal evidence in support of the hypotheses which links urate to clinical disorders.

Our recently published MR-PheWAS analysis (phenome-wide association study incorporated with Mendelian randomisation [MR] design) on the interim release data of UK Biobank (n=120,091) provided an overview of the disease outcomes that were associated with the urate genetic risk loci (461). Our study demonstrated that serum urate level shared the same genetic risk loci with multiple disease outcomes, particularly those related to cardiovascular/metabolic diseases and autoimmune disorders (461). These findings provide a rationale for the further investigation of whether these cross-phenotype associations are causal. Although we have applied multiple methodologies to distinguish the PheWAS associations that were causal from those due to pleiotropy or genetic linkage, the use of the interim release data of UK Biobank set power limitations to our investigation and did not allow us to investigate less prevalent phenotypes. The release of the full UK Biobank GWAS genotype dataset provides a unique opportunity to validate the previous MR-PheWAS findings and to include phenotypes that were not investigated in the previous study due to insufficient number of cases and controls.

In this study, we performed an updated phenome-wide Mendelian randomisation study (PWMS) by using data from the full UK Biobank cohort. A weighted polygenic risk score (GRS), incorporating effect estimates of multiple genetic risk loci taken from the most recent and largest GWAS (151), was employed as a proxy of serum urate level. The framework of phenome was defined by using both the PheCODE schema (also used in the previous MR-PheWAS) (461) and a novel Bayesian analysis framework, termed TreeWAS: tree-structured phenotypic model (460). Any replication of previous findings and/or novel findings were further explored in this study.

6.3 Methods

6.3.1 UK Biobank data

UK Biobank is a large-scale, population-based prospective cohort study, designed to investigate the genetic and non-genetic determinants of a wide range of complex diseases and phenotypes (415). The study recruited over 500,000 participants aged between 40-69 years in 2006-2010 and combined extensive measurement of baseline data and genotype data with linked national medical records (e.g. in-patient hospital episode records, cancer registry and death registry) for longitudinal follow-up. UK Biobank obtained ethical approval from the North West Multi-Centre Research Ethics Committee (11/NW/0382). The research protocol of this study was reviewed by the UK Biobank committee to ensure it was in concordance with the Ethics and Governance Framework of UK Biobank. The genotype and phenotype data used in this study were obtained from UK Biobank under an approved data request application (application ID: 10775).

Genotype data - Genotyping, quality control and genotype imputation were conducted by the UK Biobank team prior to the data release and the detailed procedures are described by Bycroft *et al* (421). The initial 50,000 participants were genotyped by the Affymetrix UK BiLEVE Axiom array and the remaining 450,000 participants were genotyped by the Affymetrix UK Biobank Axiom array. Genotype imputation was performed based on a merged reference panel of the Haplotype Reference Consortium (HRC) (426) and the UK10K haplotype resources (424), and the classical allelic variations at the MHC region were further imputed by using an additional multi-population reference panel (427). For quality control, a list of field variables was made available by the UK Biobank to indicate the genotype quality, population structure, and genetic relatedness.

Phenotype data - A variety of national health systems and sources were used by the UK Biobank to follow up the disease diagnosis, cancer occurrence, and causes of death among

the enrolled participants. Currently, there are three main different types of health records (i.e. hospital inpatient episodes, cancer registry data and death registry data) that have been incorporated into the central database. The coding for clinical diagnoses in these datasets followed the World Health Organisation's International Classification of Diseases (ICD) coding systems but used different ICD versions (ICD-10 or ICD-9) according to the date of record. Primary and/or secondary ICD codes are available in the hospital inpatient data and/or death registry data to classify the main causes and contributory causes of the event of hospitalisation and/or death respectively.

6.3.2 Study population and quality control

In order to minimise the influence of the diverse population structure in UK Biobank, our study was constrained to a subset of unrelated White British subjects with high quality genotype data. The metrics used for genotype quality control (QC) were based on the data fields created by the UK Biobank. Samples that were identified as a sex mismatch, outliers with high heterozygosity or with high missing rate, putative sex chromosome aneuploidy, individuals with excess relatives, or non-White British ancestry were all excluded from the analysis. The largest possible subset (vertices) of individuals without relatedness were identified using an algorithm implemented in the R package "*i-graph (v1.0.1)*" developed by Bycroft *et al* (421). The detailed procedures for QC and the selection of target population are described in the Supplementary Methods. As a result, a subset of 339,256 unrelated individuals of White British ancestry were finally included in analysis.

6.3.3 Weighted genetic risk score

To generate a genetic proxy for SUA level, genetic risk loci associated with SUA level were searched across the GWAS catalogue and literature. Thirty genetic variants that were identified in previous GWASs and associated with SUA at $p<5\times10^{-8}$ among European population were used in the GRS (150, 151). In comparison to the genetic instruments used in the MR analysis performed by White *et al* (323), one additional SNP, rs164009 located in the *PRPSAP1* gene ($p=7.06\times10^{-7}$), was included on the basis of its functional role in urate metabolism (encoding a protein involved in the regulation of purine synthesis). Therefore, a total of 31 independent SNPs were selected as components of the genetic proxy for SUA level. The overall proportion of variance (adjusted R^2) of SUA level explained by the 31 genetic variants was estimated to be 7% (151). The SNP effect on SUA level (effect size and standard error [SE]) was taken from the largest meta-analysis of GWAS in the European population performed by the Global Urate Genetics Consortium (GUGC) consortium (151).

Genotypes of the 31 selected SNPs were extracted from the UK Biobank genetic datasets for the target population (n=339,256). A weighted genetic risk score (GRS) was constructed by incorporating effect estimates of the 31 SUA genetic risk loci. Specifically, the polygenic risk score was created by adding up the number of SUA-increasing alleles for each SNP weighted based on the SNP effect size (regression beta coefficients) (151) and then adding this weighted score for all 31 SNPs. For instance, if an individual i carries g_{ik} copies of the SUA-increasing allele for each variant $k = 1, \ldots, 31$, the weight for variant k is w_k then their weighted polygenic score is $Z_i = \sum_{k=1}^{31} w_k g_{ik}$. The weighted polygenic risk score was calculated by using plink 2.0.

6.3.4 Phenome framework

We analysed the three phenotypic datasets (in-patient hospital records, cancer registry data, and death registry data) available in the UK Biobank database. As we were interested in disease phenotypes, the framework of the phenome was defined based on the ICD codes in the electronic medical records. We pooled the hospital episode data, cancer registry data and death registry data together and included both the primary and secondary ICD codes. The breadth of ICD-10/9 codes used in the UK Biobank well described the range of disease of participants, however, individual ICD codes could not be directly used to define the phenome, as they were designed to represent increasingly specific sub-phenotypes instead of independent phenotypes. To account for the correlations between ICD codes, we applied two strategies: (i) the PheCODE schema that has been recently updated and successfully adopted in our previous MR-PheWAS (461); and (ii) a novel Bayesian analysis framework (TreeWAS) that was developed by the researchers from the Wellcome Trust Centre for Human Genetics (460).

PheCODE schema - The PheCODE system was developed to combine one or more related ICD codes into distinct disease groups (433). To develop a phenotyping method applicable to the ICD-10 coding system in UK Biobank, we created a map to match ICD-10 codes to phecodes (461), The latest version of the PheCODE system includes 1,866 hierarchical phenotype codes that could be directly matched to the ICD-9/10 codes and provides a scheme to automatically exclude the patients that have similar or potentially overlapping disease states from the corresponding control group (e.g., excluding type 1 diabetes from being in control group when analysing the phenotype of type 2 diabetes). More details about the updated PheCODE system are described in the previous publication (461).

Tree-structured phenotypic model - A novel Bayesian analysis framework has recently been developed to interrogate the increasingly specific sub-phenotypes defined by ICD-10 coding system with increased statistical power to detect genotype-phenotype associations (460). In principle, this phenotyping method models the genetic coefficients across all phenotypes as a set of random variables. To model the correlations of the hierarchical tree-like structure of ICD-10 codes (termed as tree-structured phenotypic model]), a Markov process was applied to allow the genetic coefficients to evolve down the tree trunk and branches. The tree structure was determined based on the classification hierarchy of ICD-10 coding system, where each node in the tree represents a clinical term in the classification. More details about the tree-structured phenotyping process are described elsewhere (460).

6.3.5 Statistical analysis

To take advantage of both phenotyping models, we explored the association between the weighted GRS of urate and the phenome framework defined by both the PheCODE schema (described as PheWAS analysis) and the tree-structured phenotypic model (described as TreeWAS analysis), respectively. The correlation with weighted GRS was examined for a number of potential confounding factors including sex, age, BMI, assessment center and the first 5 PCs (**Supplementary Table 6-1**). In the PheWAS analysis, the associations between weighted GRS and phecodes (with no less than 20 cases) were examined by logistic regression. Given that many phecodes were not independent, we applied the false discovery rate (FDR) method to correct the significance threshold (corresponding to a FDR of $q<0.05$) to account for the multiple testing (473). In the TreeWAS analysis, associations between the weighted GRS and the phenome variables were tested by the Bayesian network analysis at both terminal and internal nodes of the tree structure. The marginal posterior probability (PP) for each node in the tree (where its genetic coefficient was non-zero) and the corresponding maximum posteriori effect estimate with 95% credible interval were determined by using the maximum a posteriori (MAP) estimator. Any association with any node of the tree at the PP $\geqslant 0.95$ was highlighted for further investigation. Details about the TreeWAS analysis have been described before (460). All the statistical analyses were implemented by R 3.3.2.

6.3.6 Replication in MR-base database

To validate findings, PheWAS associations were further examined in the MR-base database for replication in different populations. MR-base is a database and analytical platform for MR methods developed by the Medical Research Council, Integrative Epidemiology Unit at the University of Bristol (524). We applied this platform to replicate the findings by two-

sample MR analyses (i.e., inverse variance weighted MR [IVW MR] and Egger-MR) using summary data from the largest available GWASs for the disease outcomes of interest.

6.3.7 Sensitivity analysis

We then performed sensitivity analyses to explore any causal effect and pleiotropic association. To identify genetic variants with pleiotropy, we examined their association with a set of metabolic traits (i.e., body mass index [BMI], waist to hip ratio [WHR], total cholesterol [TC], low-density lipoprotein cholesterol [LDL-c], high-density lipoprotein cholesterol [HDL-c], apolipoprotein-A/B, fasting glucose, 2hr glucose, glycoproteins, systolic blood pressure [SBP], and diastolic blood pressure [DBP]) through the publicly available resources from various GWAS consortia: GIANT (Genetic Investigation of ANthropometric Traits) (525), GLGC (Global Lipids Genetic Consortium) (526), MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium) (527), and ICBP (International Consortium for Blood Pressure) (528). Pleiotropy was declared when these GWAS summary data reported genetic association between serum urate risk loci and these metabolic traits at $p<1.61\times10^{-3}$ (0.05/31) (**Supplementary Table 6-2, 6-3, 6-4, 6-5, 6-6**): (i) urate-specific loci: including 14 SNPs with no pleiotropic effect on the examined metabolic traits (**Supplementary Table 6-2**); (ii) urate-obesity pleiotropic loci: including 10 SNPs with pleiotropic effects on BMI or WHR (**Supplementary Table 6-3**); (iii) urate-BP pleiotropic loci: including 10 SNPs with pleiotropic effects on blood pressures (i.e., DBP and SBP) (**Supplementary Table 6-4**); (iv) urate-lipid pleiotropic loci: including 6 SNPs with pleiotropic effects on lipids (i.e., TC, LDL-c, HDL-c, apolipoprotein-A/B) (**Supplementary Table 6-5**); (v) urate-glucose pleiotropic loci: including 6 SNPs with pleiotropic effects on blood glucose (fasting glucose, 2hr glucose, glycoproteins) (**Supplementary Table 6-6**). A set of GRSs were created accordingly to re-calculate the effect estimates in PheWAS analysis.

6.4 Results

We included 339,256 unrelated White British individuals from the full UK Biobank cohort, consisting of 157,146 men and 182,110 women. The mean age of study population was 56.87 (standard deviation [SD]: 7.99) and the mean BMI was 27.40 (SD: 4.76) kg/m^2 at the time of recruitment. Other sociodemographic characteristics of the study population are summarised in **Supplementary Table 6-1**. The mean value of weighted GRS among the study population was 0.44 (SD: 0.31), which is equivalent to 0.44 mg/dL of serum urate level. The correlations between the weighted GRS and potential confounding factors (i.e.,

age, sex, BMI, assessment centre and the PCs) are examined in **Supplementary Table 6-1**. Of these, two variables (i.e., assessment centre and the PCs) were statistically significantly correlated with the weighted GRS and therefore were adjusted as covariates.

6.4.1 PheWAS and TreeWAS associations

Within the study population, we identified 10,750 unique ICD-10 codes and 3,113 ICD-9 codes in total. After mapping the diagnostic ICD-10/9 codes in UK Biobank to phecodes, the phenome defined by PheCODE schema consisted of 1807 distinct phecodes among the study population. After filtering the phecodes with no less than 20 cases, PheWAS analysis was performed for 1,431 phecodes (median number of cases: 345 [range: 20-107,298]) which could be classified into 17 broadly related disease categories (**Table 6-1**). Associations with the weighted GRS of urate were examined for 1,431 case-control groups, leading to an adjusted significance threshold of $p<3.35\times10^{-4}$ (corresponding to a FDR of $q<0.05$) to account for multiple testing. Of these, 13 phecodes were identified to be associated with genetically determined high serum urate level at $p<3.35\times10^{-4}$ (**Table 6-2**). These phecodes represent 4 disease groups: inflammatory polyarthropathies (OR=1.28; 95%CI: 1.21 to 1.35; $p=4.97\times10^{-19}$), hypertensive disease (OR=1.08; 95%CI: 1.05-1.11; $p=6.02\times10^{-7}$), circulatory disease (OR=1.05; 95%CI: 1.02 to 1.07; $p=3.29\times10^{-4}$) and metabolic disorders (OR=1.07; 95%CI: 1.03 to 1.11; $p= 3.33\times10^{-4}$), and 9 disease outcomes: gout (OR=5.37; 95%CI: 4.67 to 6.18; $p= 4.27\times10^{-123}$), gouty arthropathy (OR=5.11; 95%CI: 2.45 to 10.66; $p=1.39\times10^{-5}$), pyogenic arthritis (OR=2.10; 95%CI: 1.41 to 3.14; $p=2.87\times10^{-4}$), essential hypertension (OR=1.08; 95%CI: 1.05 to 1.11; $p=6.62\times10^{-7}$), coronary atherosclerosis (OR=1.10; 95%CI: 1.05 to 1.15; $p=1.17\times10^{-5}$), ischaemic heart disease (OR=1.10, 95%CI: 1.05 to 1.15; $p=1.73\times10^{-5}$), chronic ischaemic heart disease (OR=1.10, 95%CI: 1.05 to 1.15; $p=1.52\times10^{-5}$), myocardial infarction (OR=1.15, 95%CI:1.07 to 1.23, $p=5.23\times10^{-5}$), and hypercholesterolaemia (OR=1.08, 95%CI: 1.04 to 1.13, $p=3.34\times10^{-4}$).

Table 6 - 1: The number of phenotypes and cases in each disease category.

Disease categories	Number of phenotypes	Number of cases		
		Median	Mean	Maxium
Circulatory system	140	434	3,581	107,298
Congenital anomalies	45	102	230	1,480
Dermatological diseases	74	283	2,544	89,976
Diseases in sense organs	104	253	1,228	31,845
Digestive diseases	143	551	3,123	62,862
Neoplasms	129	493	2,558	84,098
Infectious diseases	48	190	958	8,600
Endocrine and metabolic diseases	103	154	1,590	35,954
Haematopoietic diseases	40	228	1,200	10,095
Neurological diseases	69	224	1,180	32,194
Respiratory diseases	71	674	2,448	49,782
Mental disorders	64	260	1,493	23,226
Genitourinary diseases	140	655	2,536	82,964
Pregnancy complications	28	237	914	7,518
Musculoskeletal diseases	109	347	2,847	59,852
Clinical symptoms	27	711	3,741	33,553
Injuries and poisonings	97	388	1,079	13,303

Table 6 - 2: Phenotypes associated with the weighted GRS of SUA level in PheWAS analysis ($p<3.35\times10^{-4}$).

Phecode	Description	Group	n_cases	n_controls	beta	se	OR (95%CI)	*P*-value
274.1	Gout	endocrine/metabolic	2,532	335,108	1.682	0.071	5.37 (4.67, 6.18)	4.27E-123
714	Inflammatory polyarthropathies	musculoskeletal	15,408	320,862	0.244	0.027	1.27 (1.21, 1.34)	4.97E-19
401	Hypertension	circulatory system	63,694	274,477	0.076	0.015	1.07 (1.05, 1.11)	6.02E-07
401.1	Essential hypertension	circulatory system	63,442	274,477	0.077	0.015	1.08 (1.05, 1.11)	6.26E-07
411.4	Coronary atherosclerosis	circulatory system	25,795	311,554	0.096	0.022	1.10 (1.05, 1.14)	1.17E-05
274.11	Gouty arthropathy	endocrine/metabolic	88	335,108	1.631	0.375	5.10 (2.45, 10.66)	1.39E-05
411.8	Chronic ischaemic heart disease, unspecified	circulatory system	25,567	311,554	0.095	0.022	1.09 (1.05, 1.14)	1.52E-05
411	Ischaemic Heart Disease	circulatory system	25,617	311,554	0.094	0.022	1.09 (1.05, 1.14)	1.73E-05
411.2	Myocardial infarction	circulatory system	9,829	311,554	0.138	0.034	1.14 (1.07, 1.22)	5.23E-05
711.1	Pyogenic arthritis	musculoskeletal	270	277,590	0.742	0.205	2.10 (1.41, 3.13)	2.87E-04
459.9	Circulatory disease	circulatory system	107,298	230,622	0.046	0.013	1.04 (1.02, 1.07)	3.29E-04
277	Disorders of metabolism	endocrine/metabolic	35,954	302,209	0.067	0.019	1.07 (1.03, 1.11)	3.33E-04
272.11	Hypercholesterolaemia	endocrine/metabolic	27,040	308,948	0.077	0.021	1.08 (1.04, 1.12)	3.34E-04

In the Bayesian analysis framework, containing 10,750 diagnostic terms, a total of 27 parent/child nodes of ICD-10 terms were identified with PP≥0.95. They were clustered mainly in five branches of the hierarchical tree structure (**Supplementary Table 6-7**): (i) block M10 (gout, OR=5.16, 95%CI 4.55 to 5.84; PP=1.00) and its sub-phenotypes M10.0 (idiopathic gout) and M10.9 (gout, unspecified); (ii) block I10-I15 (hypertensive disease, OR=1.07, 95%CI 1.06 to 1.08; PP>0.99) and its sub-phenotype I10 (essential hypertension); (iii) block I20-I25 (ischaemic heart diseases, OR=1.07, 95%CI 1.06 to 1.08; PP>0.99) and its sub-phenotypes I20 (angina pectoris), I21 (acute myocardial infarction), I25 (chronic ischaemic heart disease), I25.1 (atherosclerotic heart disease), I25.2 (old myocardial infarction); (iv) block I30-I52 (other forms of heart disease, OR=1.07, 95%CI 1.06 to 1.08; PP>0.99) and its sub-phenotype I50 (heart failure) and I50.1 (left ventricular failure); (v) block I60-I69 (cerebrovascular diseases, OR=1.07, 95%CI 1.06 to 1.08; PP>0.99) and its sub-phenotype I10 (cerebral infarction).

Findings from PheWAS and TreeWAS were generally consistent in their associations with gout, hypertensive disease, and heart diseases, while more sub-phenotypes were identified by TreeWAS. Association with the disease group of inflammatory polyarthropathies was statistically significant in PheWAS (OR=1.28, 95%CI: 1.21 to 1.35, $p=4.97\times10^{-19}$) but had a moderate PP in TreeWAS (OR=1.07, 95%CI: 1.06 to 1.08, PP=0.76). We examined the specific diseases included in this disease group (M05-M06: rheumatoid arthritis [RA], M07: psoriatic and enteropathic arthropathies, M08-09: juvenile arthritis, M10: gout, and M11-14: arthropathies and other arthritis), and only gout had a statistically significant association with the genetically determined serum urate levels. Association with cerebrovascular diseases had a high PP in TreeWAS (OR=1.07, 95%CI: 1.06 to 1.08, PP>0.99) but did not reach significance threshold of PheWAS (OR=1.08, 95%CI: 0.99 to 1.16, p=0.07), although their estimates were of the same direction. We re-calculated the PheWAS estimates by adding up self-reported stroke cases to increase statistical power (n=4,541), but the corresponding estimates were still not statistically significant (OR=1.05, 95%CI: 0.99 to 1.13, p=0.13, n=9,528).

6.4.2 Replication in MR-base database

To validate the findings, we performed two-sample MR analyses on associated diseases (i.e., gout, RA, CHD, myocardial infarction, ischaemic stroke) or on their corresponding intermediate traits or surrogate outcomes (i.e., SBP, DBP, total cholesterol, LDL-c, HDL-c) (**Table 6-3**). Results from IVW MR suggested that genetically determined high serum urate level was associated with increased risk of gout (OR=4.53, 95%CI: 3.64 to 5.64, $p= 9.66\times10^{-}$

[42]), CHD (OR=4.53, 95%CI: 3.64 to 5.64, P=9.66×10^{-42}), myocardial infarction (OR=4.53, 95%CI: 3.64 to 5.64, P=9.66×10^{-42}) and decreased level of HDL-c (OR=4.53, 95%CI: 3.64 to 5.64, p=9.66×10^{-42}), but had no effect on RA (OR=0.92, 95%CI: 0.84 to 1.01, p=0.085) and ischaemic stroke (OR=1.03, 95%CI: 0.93 to 1.14, p=0.582). Egger MR indicated pleiotropic effects on the causal estimates of DBP ($P_{pleiotropy} = 0.014$), SBP ($P_{pleiotropy} = 0.003$), CHD ($P_{pleiotropy} = 0.008$), myocardial infarction ($P_{pleiotropy} = 0.014$) and HDL-c ($P_{pleiotropy} = 0.014$). When balancing out the potential pleiotropic effects in Egger MR, causal effect can only be verified for gout (OR=4.17, 95%CI: 3.03 to 5.74, $P_{effect} = 1.27 \times 10^{-9}$; $P_{pleiotropy} = 0.485$).

Table 6 - 3: Replication of MR effect estimates in MR-base database.

Outcome	beta	se	OR (95% CI)	P *effect*	P *pleiotropy*	n_cases	n_total	Data source
Replication of significant PheWAS findings								
Gout								
PheWAS	1.682	0.071	5.37 (4.67, 6.18)	4.27E-123	--	2,532	337,640	UKBB
IVW MR	1.511	0.112	4.53 (3.64, 5.64)	9.66E-42	--	2,115	67,259	GUGC
Egger MR	1.427	0.163	4.17 (3.03, 5.74)	1.27E-09	0.485			
Hypertension								
PheWAS	0.076	0.015	1.07 (1.05, 1.11)	6.02E-07	--	63,694	338,171	UKBB
DBP								
IVW MR	0.427	0.272	1.53 (0.90, 2.61)	0.116	--	--	69,395	ICBP
Egger MR	-0.219	0.351	0.80 (0.40, 1.60)	0.538	0.014	--		
SBP								
IVW MR	0.409	0.402	1.51 (0.68, 3.31)	0.308	--	--	69,395	ICBP
Egger MR	-0.713	0.496	0.49 (0.19, 1.29)	0.161	0.003	--		
Coronary heart disease								
PheWAS	0.094	0.022	1.09 (1.05, 1.14)	1.73E-05	--	25,617	337,171	UKBB
IVW MR	0.098	0.038	1.10 (1.02, 1.19)	0.009	--	60,801	123,504	CARDIoGRAMplusC4D
Egger MR	0.001	0.048	1.00 (0.91, 1.10)	0.977	0.008			
Myocardial infarction								
PheWAS	0.138	0.034	1.14 (1.07, 1.22)	5.23E-05	--	9,829	321,383	UKBB
IVW MR	0.105	0.041	1.00 (0.90, 1.11)	0.983	--	43,676	128,199	CARDIoGRAMplusC4D
Egger MR	-0.001	0.053	1.11 (1.02, 1.20)	0.011	0.008			

Hypercholesterolaemia								
PheWAS	0.077	0.021	1.08 (1.04, 1.12)	3.34E-04	--	27,040	335,988	UKBB
Total cholesterol								
IVW MR	0.028	0.036	1.03 (0.96, 1.10)	0.433	--	--	94,595	GLGC
Egger MR	0.048	0.052	1.05 (0.95, 1.16)	0.368	0.602	--		
HDL-c								
IVW MR	-0.075	0.026	0.93 (0.88, 0.98)	4.00E-03	--	--	94,311	GLGC
Egger MR	-0.010	0.035	0.99 (0.92, 1.06)	0.767	0.016	--		
LDL-c								
IVW MR	0.011	0.023	1.05 (0.95, 1.16)	0.623	--	--	89,888	GLGC
Egger MR	0.045	0.033	1.05 (0.98, 1.12)	0.188	0.175	--		
Replication of non-significant PWMR findings								
RA								
PheWAS	0.095	0.055	1.10 (0.99, 1.22)	0.683	--	3,522	324,384	UKBB
IVW MR	-0.081	0.047	0.92 (0.84, 1.01)	0.085	--	19,234	61,565	MR-base
Egger MR	-0.103	0.066	0.90 (0.79, 1.03)	0.132	0.645			
Ischaemic stroke								
PheWAS	0.071	0.04	1.08 (0.99-1.16)	0.070	--	9,528	338,172	UKBB
IVW MR	0.029	0.052	1.03 (0.93, 1.14)	0.582	--	10,307	19,326	ISGC
Egger MR	-0.028	0.074	0.97 (0.84, 1.12)	0.707	0.290			

UKBB, UK Biobank; GUGC, Global Urate Genetics Consortium; ICBP, International Consortium of Blood Pressure; CARDIoGRAMplusC4D, Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) Genetics) consortium; GLGC, Global Lipids Genetics Consortium; ISGC, International Stroke Genetics Consortium.

6.4.3 Sensitivity analysis

To further investigate the influence of pleiotropy, we re-calculated the PheWAS estimates by using a number of GRSs created based on their association with a set of metabolic traits (**Figure 6-1**). The GRS of urate-specific loci was only associated with gout and its encompassing disease group of inflammatory polyarthropathies, but not with any cardiovascular/metabolic diseases. In contrast, the GRSs of loci pleiotropic influence on obesity, BP, lipids and glucose showed significant association with both gout and the cardiovascular/metabolic diseases. Specifically, the GRS of pleiotropic loci on lipids was significantly associated with all cardiovascular/metabolic diseases, including hypertensive diseases (i.e., essential hypertension), heart diseases (i.e., ischaemic heart diseases), and metabolic disorders (i.e., hypercholesterolaemia). Additionally, the GRS of pleiotropic loci on glucose was significantly associated with diabetes mellitus (i.e., type 2 diabetes). When removing any group of pleiotropic loci from the creation of GRS, their association with hypertensive diseases, heart diseases, and metabolic disorders were not statistically significant (**Table 6-4**). The effects of pleiotropic loci (mapped with genes) on SUA levels against their effects on four representative disease outcomes were plotted in **Figure 6-2**, in which the two urate transporter genes (*SLC2A9* and *ABCG2*) are recognised as the leading loci driving the association with gout, the *GCKR* gene is the leading locus driving the association with hypercholesterolaemia, and the *PTPN11/ATXN2* gene is the leading locus driving the association with hypertension and ischaemic heart diseases.

Figure 6 - 1: A network plot for sensitivity analysis of PheWAS using different sets of weighted GRSs.

(M05-M14: Inflammatory polyarthropathies; I10-I15: Hypertensive diseases; I20-I25: Ischaemic heart diseases; E70-E90: Metabolic disorders)

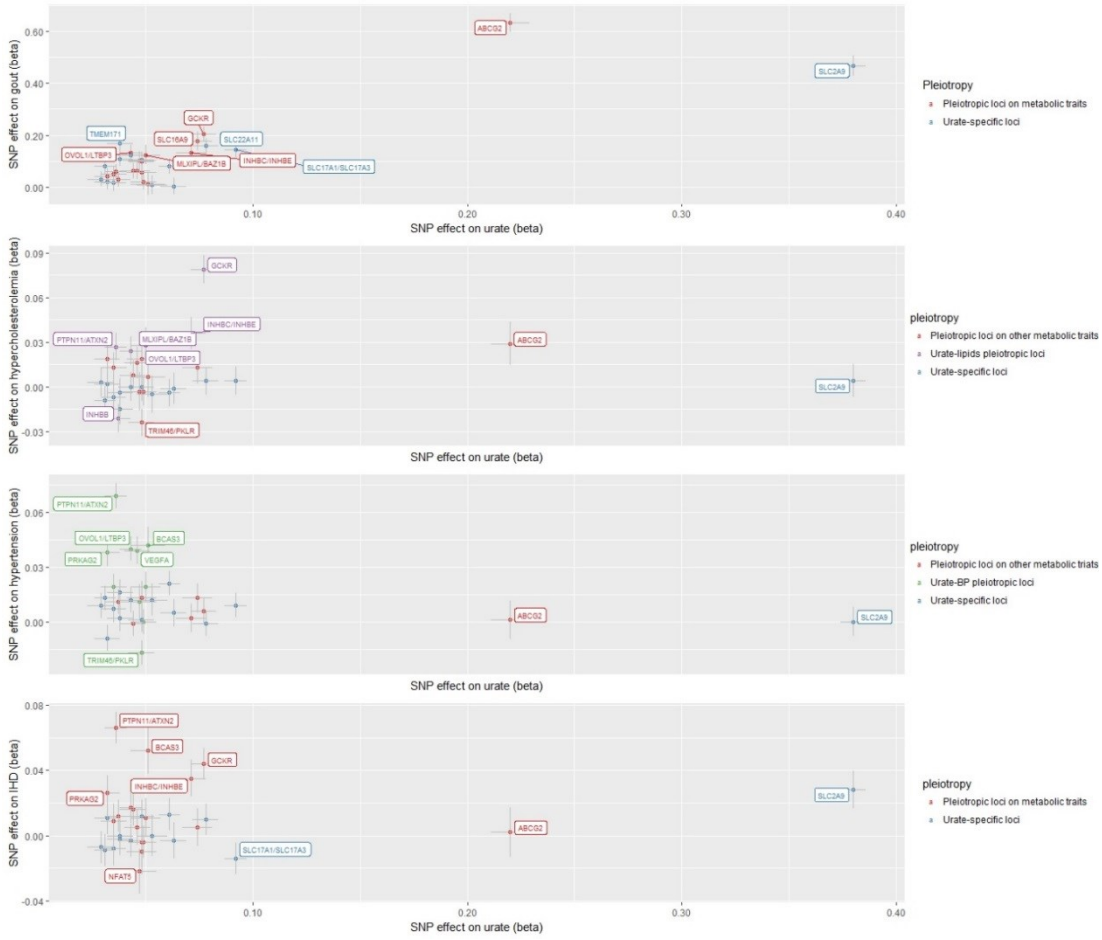


Figure 6 - 2: A scatter plot for the effects of pleiotropic loci (mapped with genes) on SUA levels against their effects on four representative disease outcomes.

Table 6 - 4: Sensitivity analysis by removing the pleiotropic loci on metabolic traits.

Disease outcomes	GRS of all-urate loci (n=31)			GRS of loci without pleiotropy on obesity (n=21)			GRS of loci without pleiotropy on BP (n=21)			GRS of loci without pleiotropy on lipids (n=25)			GRS of loci without pleiotropy on glucose (n=28)		
	OR (95%CI)	*pval*	*FDR*	*OR (95%CI)*	*pval*	*FDR*	*OR (95%CI)*	*pval*	*FDR*	*OR (95%CI)*	*pval*	*FDR*	*OR (95%CI)*	*pval*	*FDR*
Gout	5.37 (4.67, 6.18)	4.27E-123	TRUE	3.89 (3.32, 4.56)	1.01E-62	TRUE	5.42 (4.68, 6.28)	6.00E-113	TRUE	5.07 (4.38, 5.86)	5.27E-105	TRUE	5.09 (4.41, 5.88)	1.31E-108	TRUE
Inflammatory polyarthropathies	1.27 (1.21, 1.34)	4.97E-19	TRUE	1.21 (1.14, 1.28)	6.43E-10	TRUE	1.26 (1.19, 1.33)	1.02E-15	TRUE	1.26 (1.19, 1.33)	1.16E-15	TRUE	1.27 (1.20, 1.34)	1.34E-17	TRUE
Hypertension	1.07 (1.05, 1.11)	6.02E-07	TRUE	1.07 (1.03, 1.10)	2.31E-04	FALSE	1.04 (1.01, 1.07)	2.24E-02	FALSE	1.05 (1.02, 1.09)	0.002	FALSE	1.08 (1.05, 1.11)	1.82E-06	TRUE
Essential hypertension	1.08 (1.05, 1.11)	6.26E-07	TRUE	1.07 (1.03, 1.10)	2.44E-04	FALSE	1.04 (1.01, 1.07)	2.23E-02	FALSE	1.05 (1.02, 1.09)	0.002	FALSE	1.08 (1.05, 1.11)	1.83E-06	TRUE
Coronary atherosclerosis	1.10 (1.05, 1.14)	1.17E-05	TRUE	1.08 (1.03, 1.13)	0.001	FALSE	1.08 (1.03, 1.13)	7.82E-04	FALSE	1.05 (1.01, 1.10)	0.022	FALSE	1.08 (1.03, 1.13)	4.88E-04	FALSE
Gouty arthropathy	5.10 (2.45, 10.66)	1.39E-05	TRUE	4.04 (1.73, 9.44)	0.001	FALSE	5.69 (2.61, 12.37)	1.19E-05	TRUE	5.36 (2.47, 11.63)	2.16E-05	TRUE	4.78 (2.24, 10.21)	5.37E-05	TRUE
Chronic ischaemic heart disease	1.09 (1.05, 1.14)	1.52E-05	TRUE	1.08 (1.03, 1.13)	0.002	FALSE	1.08 (1.03, 1.13)	8.61E-04	FALSE	1.05 (1.01, 1.10)	0.024	FALSE	1.08 (1.03, 1.13)	5.58E-04	FALSE
Ischaemic Heart Disease	1.09 (1.05, 1.14)	1.73E-05	TRUE	1.08 (1.03, 1.13)	0.002	FALSE	1.08 (1.03, 1.13)	9.51E-04	FALSE	1.05 (1.01, 1.10)	0.026	FALSE	1.08 (1.03, 1.13)	6.04E-04	FALSE
Myocardial infarction	1.14 (1.07, 1.22)	5.23E-05	TRUE	1.11 (1.03, 1.20)	0.006	FALSE	1.11 (1.04, 1.19)	0.003	FALSE	1.08 (1.01, 1.16)	0.033	FALSE	1.12 (1.04, 1.20)	0.002	FALSE
Pyogenic arthritis	2.10 (1.41, 3.13)	2.87E-04	TRUE	1.96 (1.24, 3.09)	0.004	FALSE	2.14 (1.41, 3.26)	3.73E-04	FALSE	1.82 (1.20, 2.77)	0.005	FALSE	1.95 (1.29, 2.95)	0.001	FALSE
Circulatory disease	1.04 (1.02, 1.07)	3.29E-04	TRUE	1.03 (1.01, 1.06)	0.020	FALSE	1.03 (1.00, 1.05)	0.048	FALSE	1.03 (1.00, 1.05)	0.041	FALSE	1.04 (1.01, 1.07)	0.003	FALSE
Disorders of metabolism	1.07 (1.03, 1.11)	3.33E-04	TRUE	1.04 (1.00, 1.09)	0.038	FALSE	1.07 (1.03, 1.11)	0.001	FALSE	1.03 (0.99, 1.07)	0.094	FALSE	1.05 (1.01, 1.09)	0.019	FALSE
Hypercholesterolaemia	1.08 (1.04, 1.12)	3.34E-04	TRUE	1.03 (0.98, 1.07)	0.293	FALSE	1.07 (1.02, 1.12)	0.003	FALSE	1.03 (0.98, 1.07)	0.233	FALSE	1.04 (1.00, 1.09)	0.051	FALSE

6.5 Discussion

The present study demonstrated that genetically determined high serum urate level was consistently associated with increased risk of several disease groups, including inflammatory polyarthropathies (e.g., gout and gouty arthropathy), hypertensive diseases (e.g., essential hypertension), heart diseases (e.g., coronary atherosclerosis, myocardial infarction, angina pectoris, ischaemic heart disease and heart failure) and metabolic disorders (e.g., hypercholesterolaemia). This study using data from the full UK Biobank cohort (n=339,256) replicated the associations discovered in the previous MR-PheWAS study based on the interim release of UK Biobank genetic data (n=120,091) (461), and identified a number of new sub-phenotypes of diseases (e.g., gouty arthropathy, angina pectoris, and heart failure). Association between urate and the risk of gout, CHD, myocardial infarction and decreased level of HDL-c were also successfully replicated in different populations by analysing various GWAS consortium data documented in the MR-base database (524), but a causal relationship was only supported for gout. The role of urate in the development of cerebrovascular diseases is debatable, as their association was only identified in TreeWAS but cannot be replicated in PheWAS or MR analysis. Overall, findings from the current study supported the epidemiological observations that high serum urate level is correlated with high risk of hypertensive diseases, heart diseases, and metabolic disorders, and indicated that their associations were likely due to genetic pleiotropy instead of causality.

6.5.1 Main findings and possible explanations

Our finding that genetically predicted serum urate level is causally associated with increased risk of gout and its sub-phenotypes is not surprising, as it is well known that the causal factor of gout is represented by the monosodium urate crystals (MSU), which leads to acute local inflammation in joints (529). Moreover, this study also detected an association between urate and the disease group of inflammatory polyarthropathies. To investigate if there was any other type of inflammatory polyarthropathies associated with urate, we examined the association of urate with all specific diseases included in this group, but none of them are statistically significant. When removing gout cases from analysis, the disease group of inflammatory polyarthropathies was no longer associated with urate, indicating the observed association was driven by gout. As emerging evidence supports the notion that urate has pro-inflammatory effects and may contribute to a growing family of auto-inflammatory disease (e.g., RA) through the production of a panel of inflammatory cytokines including IL-1β, IL-8, TNF-α, and IL-6 (529) (530). It is hypothesised that these findings are of relevance to RA, as IL-6 is a key pro-inflammatory cytokine that is originally described as a T-cell-derived B-

cell differentiation factor and may stimulate the production of autoantibodies such as rheumatoid factor (531, 532). In this study, we additionally examined the association between the GRS of urate and RA in both UK biobank population (3,522 cases) and a larger European population (19,234 cases) (533), but neither of them detected any significant association. Our study did not provide any supportive evidence on the aetiological relevance between urate and RA as suggested by prior studies.

Numerous epidemiological studies have reported that elevated serum urate level is related to increased risk of hypertension and their relationship has been consistent, showing a dose-response relationship and of similar magnitude (486). Findings from this study supported their association, but the magnitude of estimated effect size (OR=1.07; 95%CI: 1.05 to 1.11) is relatively smaller than that of traditional epidemiological studies, in which a recent meta-analysis synthesised data from 97,824 individuals and reported a pooled OR of 1.15 (95%CI: 1.06 to 1.26) for incident hypertension for a 1 mg/dL increase of serum urate level (534). Similarly, the association between urate and a multitude of cardiac events of varying severity, including coronary atherosclerosis, angina pectoris, ischaemic heart diseases, acute/old myocardial infarction and heart failure, were also well explored in this study. The PheWAS, TreeWAS, and IVW MR replication analysis consistently supported a moderate association between urate and different types of heart diseases. However, the Egger-MR analysis reported here provided no evidence for causality, but suggested the presence of pleiotropy in their associations.

Large epidemiological studies have established an association between high serum urate level and the increased risk of metabolic disorders (535). The NHANES III survey study suggested that high serum urate level was associated with increased level of serum LDL-c, triglycerides, total cholesterol, apolipoprotein-B, and decreased level of HDL-c (536). Our study further strengthened this epidemiological evidence and highlighted an association between urate and hypercholesterolaemia. Our IVW MR analysis replicated the corresponding association with its surrogate outcome (i.e., HDL-c), but the Egger-MR analysis suggested the presence of pleiotropy instead of causality. Additionally, epidemiological studies have also indicated that high serum urate level is associated with increased risk of diabetes (537). However, this association was not detected in the main PheWAS or TreeWAS analysis, while sensitivity analysis using the GRS of urate-glucose pleiotropic loci (i.e., *GCKR*, *IGF1R*, and *SLC16A9*) identified significant association with type 2 diabetes.

To explore how genetic pleiotropy influences the association with cardiovascular/metabolic diseases, we analysed all 31 urate loci across a set of metabolic traits and identified 14 SNPs (urate-specific loci) that were exclusively associated with urate and 17 SNPs (pleiotropic loci) that were associated with metabolic traits. When examining the urate-specific loci, their GRS was only associated with gout and its encompassing disease group of inflammatory polyarthropathies, but not with any cardiovascular or metabolic diseases. In contrast, when categorising the pleiotropic loci into different groups (e.g., GRS of urate-obesity loci, GRS of urate-BP loci, GRS of urate-lipids loci and GRS of urate-glucose loci), the GRSs of pleiotropic loci showed consistent associations with both gout and the cardiovascular/metabolic diseases. When removing any group of pleiotropic loci from the creation of GRS (e.g., GRS of urate without pleiotropic loci on BP, or GRS of urate without pleiotropic loci on lipids), their association with heart diseases and metabolic disorders was not statistically significant. Based on these findings, our study suggests that the association between urate and cardiovascular/metabolic diseases is probably due to the pleiotropic effects of genetic variants on urate and metabolic traits.

Examining the associations between individual urate genetic risk loci and the related disease outcomes highlighted two loci, *GCKR* and *PTPN11/ATXN2* that drive their association with hypercholesterolaemia, hypertension and ischaemic heart disease. Pathway network analysis of the leading pleiotropic genes provides some clues on how genetic pleiotropy contributes to the association between urate and cardiovascular/metabolic disease. Genetic variation in *GCKR* is shown to be associated with concentrations of urate, triglyceride and glucose (538). The most plausible explanation for this observation is that *GCKR* affects both serum urate, triglyceride and glucose levels by a common unconfirmed mediator which is proposed to be glucose-6-phosphate (539). *GCKR* controls the hepatic production of glucose-6-phosphate, which is catabolised for triglyceride synthesis via glycolysis, into pyruvate, and acetyl coenzyme A, while glucose-6-phosphate is also a precursor of purine (uric acid) metabolism (539). Additionally, gene functional annotation of *PTPN11/ATXN2* highlights another subnetwork around haemostasis pathways, including platelet activation, aggregation, and sensitisation (activated by LDL-c) (540), and these may be relevant to the observed association with hypertension and heart diseases, but how this gene influences serum urate levels has not yet been clearly demonstrated.

6.5.2 Clinical implications and future research

The detection of a multitude of cross-phenotype associations in this study adds our understanding of the extent of shared genetic/biological components between urate and

metabolic traits. Further characterising the associations between urate and disease outcomes as causal or pleiotropic contributes to our knowledge of how the role of urate should be interpreted and used in clinical practice in the management of related disease conditions. Given that the associations between urate and cardiovascular/metabolic diseases are more likely due to pleiotropy rather than causality, our study supports the notion that urate could be a predictor but not be a target for the development of compounds that could reduce cardiovascular/metabolic disease risk. The linked biological pathways between urate and metabolic traits indicated that the frequent co-existence of gout with hypertension, cardiovascular diseases and hyperlipidaemia is a range of inter-related disease outcomes due to linked pathogenic components, rather than isolated events. This supports the European League against Rheumatism (EULAR) recommendation of systematic screening and assessment of cardiovascular/metabolic comorbidities in gout patients (541). The finding of genetic pleiotropy indicates the existence of common upstream pathological elements influencing both urate and metabolic traits, and this may suggest new opportunities and challenges for developing drugs targeting a more distal mediator that would be beneficial for both the treatment of gout and the prevention of cardiovascular/metabolic comorbidities. This study has focused on the detection of cross-phenotype associations and highlighted the importance of pleiotropy in the links of these complex diseases. We have made efforts to try to understand the cross-phenotype association in the context of a pleiotropy model, but functionally characterising the underlying biological mechanisms remains a challenge in this field and is worthy of further investigation.

6.5.3 Strengths and weaknesses of the study

The strengths of this study include its potential to examine a broad spectrum of disease outcomes related to urate and to reflect the shared biological relevance among associated phenotypes. Compared to the previous MR-PheWAS (461), the present study extends the prior findings by combining genetic risk loci of urate into a weighted GRS, exploring genetic pleiotropy on a set of metabolic traits systematically, investigating more disease outcomes, assessing their associations with >3-fold more cases, examining consistency of findings across two different phenotyping models to reduce the probability of false positive/negative findings due to factors related to the model, and replicating the findings by performing two-sample MR in different populations. Our study demonstrated the performance of two phenotyping models by accounting for the differences in the specificity and granularity of different phenome definitions and by characterising the phenotypic correlations among different levels of ICD hierarchy. TreeWAS is shown to increase statistical power by up to

20%, and can detect new associations missed by conventional PheWAS. One of the major accomplishments of this study together with the previous MR-PheWAS have been the establishment of a framework or workflow for PheWAS (461). We believe this study would be an excellent starting point for researchers who plan to use the UK Biobank resource to comprehensively interrogate the clinical significance of biomarkers. The updated version of the PheCODE schema used in this study is available for researchers who are interested in performing PheWAS in UK Biobank when requested.

This study also has limitations. The causal inference in our study is limited by the common difficulty of pleiotropy caused by the use of multiple genetic instruments. Although we have performed sensitivity analyses by grouping the pleiotropic loci based on metabolic traits and exploring their association separately, there is still a probability of undetected pleiotropy or the possibility that the relatively weak causal effects of urate on diseases were concealed by the strong pleiotropic effects of the genetic variants on metabolic traits. Moreover, as most patients (cases) were identified from the in-patient hospital records, this may have impaired the coverage of case ascertainment, especially for the diseases that do not usually cause events for hospitalisation. The incorporation of self-reported data would improve this limitation but is likely to mistakenly include cases who do not have a true diagnosis and introduce information biases. As UK Biobank is currently conducting biomarker assays and processing linkage to general practice records and out-patient data, it would be beneficial to confirm the potential association based on a widely-covered and accurately-defined criteria for case ascertainment in the future.

6.6 Conclusion

Overall, when taken together the findings from PheWAS/TreeWAS, MR replication and sensitivity analysis, we conclude a robust association between urate and a group of diseases, including gout, hypertensive diseases, heart diseases and metabolic disorders of lipids, but the causal role of urate is only supported in gout. Our study indicates that the association between urate and cardiovascular/metabolic diseases is probably due to the pleiotropic effects of genetic variants on urate and metabolic traits. These findings suggest that urate could be a good predictor for the cardiovascular/metabolic disease risk. Further investigation on therapies targeting on the shared biological pathways between urate and metabolic traits would be beneficial for the treatment of gout and the primary prevention of cardiovascular/metabolic comorbidities.

6.7 Supplementary information

Supplementary Table 6 - 1: Association between the GRS of urate and potential confounding factors.

Continuous variable	**Mean (SD)**	**Beta (SE)**	***P*-value**
Age	56.87 (7.99)	0.010 (0.044)	0.830
BMI	27.40 (4.76)	-0.023 (0.027)	0.381
PC1 score	-12.35 (1.61)	0.007 (0.009)	0.408
PC2 score	3.78 (1.50)	-0.023 (0.008)	0.007*
PC3 score	-1.59 (1.58)	-0.003 (0.009)	0.753
PC4 score	1.29 (2.94)	0.104 (0.016)	1.74e-10*
PC5 score	-0.81 (6.61)	0.344 (0.037)	2.20e-16*
Categorical variable	**Levels**	**F-value**	***P*-value**
Sex	male/female	0.476	0.490
Assessment centre	22 centres	3.451	1.41e-07*

BMI, body mass index; PC, genetic principal component.

Supplementary Table 6 - 2: A summary of 31 urate SNPs identified in previous GWAS.

SNP	Chr	Closest/GRAIL gene	Effect allele	Allele freq	beta	se	P	Pleiotropy
rs10821905	10	*A1CF/ASAH2*	A	0.824	0.053	0.007	3.45E-12	No
rs1165151	6	*SLC17A1/SLC17A3*	T	0.549	-0.092	0.005	4.52E-60	No
rs12498742	4	*SLC2A9/SLC2A9*	A	0.232	0.380	0.006	0.00E+00	No
rs1394125	15	*UBE2Q2/NRG4*	A	0.638	0.043	0.006	9.78E-11	No
rs1471633	1	*PDZK1/PDZK1*	A	0.538	0.061	0.005	1.40E-26	No
rs164009	17	*QRICH2/PRPSAP1*	A	0.387	0.029	0.006	7.06E-07	No
rs17632159	5	*TMEM171/TMEM171*	C	0.697	-0.038	0.006	2.00E-09	No
rs17786744	8	*STC1/STC1*	A	0.410	-0.031	0.005	8.82E-08	No
rs2078267	11	*SLC22A11/SLC22A11*	T	0.452	-0.078	0.006	8.73E-36	No
rs675209	6	*RREB1/RREB1*	T	0.731	0.063	0.006	1.38E-21	No
rs6770152	3	*SFMBT1/MUSTN1*	T	0.424	-0.048	0.006	2.66E-16	No
rs7188445	16	*MAF/MAF*	A	0.672	-0.032	0.006	1.15E-07	No
rs7224610	17	*HLF/HLF*	A	0.396	-0.038	0.006	4.74E-11	No
rs742132	6	*LRRC16A/LRRC16A*	A	0.294	0.035	0.006	1.90E-08	No
rs10480300	7	*PRKAG2/PRKAG2*	T	0.727	0.032	0.006	9.37E-07	Yes
rs11264341	1	*TRIM46/PKLR*	T	0.571	-0.048	0.006	1.04E-14	Yes
rs1171614	10	*SLC16A9/SLC16A9*	T	0.769	-0.074	0.007	6.48E-23	Yes
rs1178977	7	*BAZ1B/MLXIPL*	A	0.198	0.050	0.007	6.68E-12	Yes
rs1260326	2	*GCKR/GCKR*	T	0.607	0.077	0.006	1.31E-40	Yes
rs17050272	2	*INHBB/INHBB*	A	0.589	0.037	0.006	9.36E-09	Yes
rs2079742	17	*BCAS3/C17orf82*	T	0.136	0.051	0.008	6.24E-09	Yes
rs2231142	4	*ABCG2/ABCG2*	T	0.887	0.220	0.009	4.43E-116	Yes
rs2307394	2	*ORC4L/ACVR2A*	T	0.303	-0.035	0.006	7.26E-09	Yes
rs2941484	8	*HNF4G/HNF4G*	T	0.553	0.049	0.006	3.91E-17	Yes
rs3741414	12	*INHBC/INHBE*	T	0.755	-0.071	0.007	9.79E-22	Yes

SNP	Chr	Closest/GRAIL gene	Effect allele	Allele freq	beta	se	p-val	Pleiotropy
rs478607	11	*NRXN2/SLC22A12*	A	0.153	-0.048	0.007	5.31E-10	Yes
rs642803	11	*OVOL1/LTBP3*	T	0.536	-0.043	0.005	4.51E-14	Yes
rs653178	12	*ATXN2/PTPN11*	T	0.483	-0.036	0.005	2.45E-10	Yes
rs6598541	15	*IGF1R/IGF1R*	A	0.645	0.044	0.006	5.20E-13	Yes
rs7193778	16	*NFAT5/NFAT5*	T	0.150	-0.047	0.008	2.36E-08	Yes
rs729761	6	*VEGFA/VEGFA*	T	0.715	-0.046	0.006	3.05E-12	Yes

Supplementary Table 6 - 3: A summary of the pleiotropic effect of urate SNPs on obesity.

SNP	Chr	Closest/GRAIL gene	effect_allele	BMI (n=336,107)			WHR (n=141,537)			Pleiotropy
				beta	se	p	beta	se	P	
rs2231142	4	*ABCG2/ABCG2*	T	-0.081	0.017	2.16E-06	0.029	0.025	0.261	Yes
rs7193778	16	*NFAT5/NFAT5*	T	0.339	0.072	2.17E-06	0.468	0.130	3.10E-04	Yes
rs2941484	8	*HNF4G/HNF4G*	T	0.207	0.050	3.07E-05	-0.204	0.088	0.020	Yes
rs1260326	2	*GCKR/GCKR*	T	-0.131	0.032	4.22E-05	0.130	0.045	0.004	Yes
rs478607	11	*NRXN2/SLC22A12*	A	-0.281	0.069	5.08E-05	0.152	0.119	0.200	Yes
rs11264341	1	*TRIM46/PKLR*	T	0.192	0.050	1.40E-04	-0.152	0.092	0.097	Yes
rs642803	11	*OVOL1/LTBP3*	T	0.198	0.056	4.22E-04	0.133	0.077	0.084	Yes
rs653178	12	*ATXN2/PTPN11*	T	-0.233	0.067	4.93E-04	0.072	0.097	0.458	Yes
rs6598541	15	*IGF1R/IGF1R*	A	0.194	0.057	6.89E-04	0.089	0.077	0.251	Yes
rs1178977	7	*BAZ1B/MLXIPL*	A	-0.179	0.060	0.003	0.360	0.086	2.84E-05	Yes
rs1471633	1	*PDZK1/PDZK1*	A	0.091	0.039	0.021	-0.069	0.072	0.340	No
rs2079742	17	*BCAS3/C17orf82*	T	-0.158	0.069	0.022	-0.039	0.129	0.762	No
rs3741414	12	*INHBC/INHBE*	T	-0.060	0.039	0.128	0.008	0.056	0.881	No
rs7224610	17	*HLF/HLF*	A	-0.099	0.065	0.128	-0.082	0.113	0.471	No
rs17050272	2	*INHBB/INHBB*	A	0.099	0.066	0.133	-0.016	0.124	0.896	No
rs729761	6	*VEGFA/VEGFA*	T	-0.087	0.058	0.136	0.261	0.104	0.012	No
rs1171614	10	*SLC16A9/SLC16A9*	T	0.057	0.039	0.138	0.080	0.078	0.309	No
rs2307394	2	*ORC4L/ACVR2A*	T	0.106	0.075	0.159	-0.031	0.129	0.807	No
rs1394125	15	*UBE2Q2/NRG4*	A	-0.076	0.058	0.193	-0.067	0.114	0.554	No
rs12498742	4	*SLC2A9/SLC2A9*	A	0.010	0.007	0.203	-0.001	0.013	0.952	No
rs17786744	8	*STC1/STC1*	A	0.098	0.079	0.217	0.155	0.139	0.264	No
rs10480300	7	*PRKAG2/PRKAG2*	T	-0.103	0.084	0.222	-0.172	0.150	0.252	No
rs675209	6	*RREB1/RREB1*	T	-0.046	0.043	0.280	-0.206	0.079	0.009	No
rs742132	6	*LRRC16A/LRRC16A*	A	0.075	0.075	0.320	0.174	0.131	0.185	No
rs1165151	6	*SLC17A1/SLC17A3*	T	-0.025	0.026	0.345	0.050	0.046	0.273	No
rs6770152	3	*SFMBT1/MUSTN1*	T	0.039	0.051	0.448	0.044	0.090	0.625	No
rs2078267	11	*SLC22A11/SLC22A11*	T	-0.023	0.031	0.453	-0.104	0.055	0.060	No

SNP	Chr	Closest/GRAIL gene	Effect allele	BMI (n=336,107)			WHR (n=141,537)			Pleiotropy
				beta	se	p	beta	se	P	
rs10821905	10	*A1CF/ASAH2*	A	-0.031	0.060	0.600	0.053	0.106	0.617	No
rs17632159	5	*TMEM171/TMEM171*	C	0.027	0.069	0.694	-0.016	0.126	0.901	No
rs7188445	16	*MAF/MAF*	A	0.011	0.080	0.893	-0.203	0.141	0.149	No
rs164009	17	*QRICH2/PRPSAP1*	A	-0.007	0.085	0.936	-0.100	0.152	0.510	No

Supplementary Table 6 - 4: A summary of the pleiotropic effect of urate SNPs on blood pressure.

SNP	Chr	Closest/GRAIL gene	Effect allele	DBP (n=317,754)			SBP (n=317,756)			Pleiotropy
				beta	se	P	beta	se	P	
rs653178	12	*ATXN2/PTPN11*	T	1.057	0.068	6.66E-54	0.585	0.068	1.16E-17	Yes
rs642803	11	*OVOL1/LTBP3*	T	0.377	0.057	4.83E-11	0.269	0.057	2.85E-06	Yes
rs2307394	2	*ORC4L/ACVR2A*	T	0.396	0.077	2.47E-07	0.091	0.077	0.235	Yes
rs10480300	7	*PRKAG2/PRKAG2*	T	0.364	0.086	2.38E-05	0.453	0.086	1.45E-07	Yes
rs729761	6	*VEGFA/VEGFA*	T	0.236	0.060	7.38E-05	-0.040	0.060	0.506	Yes
rs2941484	8	*HNF4G/HNF4G*	T	0.194	0.051	1.30E-04	0.155	0.051	0.002	Yes
rs1178977	7	*BAZ1B/MLXIPL*	A	0.230	0.062	1.93E-04	0.026	0.062	0.676	Yes
rs7193778	16	*TRIM46/PKLR*	T	-0.081	0.073	0.268	0.347	0.073	2.19E-06	Yes
rs11264341	1	*BCAS3/C17orf82*	T	0.146	0.052	0.005	0.173	0.052	8.20E-04	Yes
rs2079742	17	*NFAT5/NFAT5*	T	0.136	0.070	0.054	0.257	0.070	2.52E-04	Yes
rs6770152	3	*SFMBT1/MUSTN1*	T	0.154	0.052	0.003	0.119	0.052	0.022	No
rs7188445	16	*MAF/MAF*	A	-0.214	0.082	0.009	-0.037	0.082	0.655	No
rs7224610	17	*HLF/HLF*	A	0.174	0.067	0.009	0.202	0.067	0.002	No
rs12498742	4	*SLC2A9/SLC2A9*	A	0.017	0.008	0.023	0.009	0.008	0.219	No
rs1471633	1	*PDZK1/PDZK1*	A	0.080	0.040	0.048	0.041	0.040	0.313	No
rs17786744	8	*STC1/STC1*	A	0.140	0.081	0.083	-0.114	0.081	0.159	No
rs17050272	2	*INHBB/INHBB*	A	0.111	0.068	0.099	0.063	0.068	0.354	No
rs3741414	12	*INHBC/INHBE*	T	0.062	0.040	0.126	0.125	0.040	0.002	No
rs1165151	6	*SLC17A1/SLC17A3*	T	0.041	0.027	0.129	0.069	0.027	0.010	No
rs1171614	10	*SLC16A9/SLC16A9*	T	0.048	0.039	0.224	0.103	0.039	0.009	No
rs17632159	5	*TMEM171/TMEM171*	C	-0.063	0.071	0.368	-0.066	0.070	0.349	No
rs10821905	10	*A1CF/ASAH2*	A	0.046	0.061	0.447	0.174	0.061	0.004	No
rs478607	11	*NRXN2/SLC22A12*	A	0.049	0.071	0.490	0.106	0.071	0.136	No
rs2231142	4	*ABCG2/ABCG2*	T	-0.012	0.018	0.497	-0.050	0.018	0.005	No
rs2078267	11	*SLC22A11/SLC22A11*	T	0.020	0.032	0.523	0.000	0.032	0.999	No
rs164009	17	*QRICH2/PRPSAP1*	A	0.050	0.087	0.570	-0.025	0.087	0.770	No
rs1394125	15	*UBE2Q2/NRG4*	A	0.026	0.060	0.658	0.042	0.060	0.477	No

SNP	Chr	Closest/GRAIL gene	Effect allele	DBP (n=317,754)			SBP (n=317,756)			Pleiotropy
				beta	se	P	beta	se	P	
rs6598541	15	*IGF1R/IGF1R*	A	0.023	0.059	0.698	-0.029	0.059	0.618	No
rs1260326	2	*GCKR/GCKR*	T	-0.012	0.033	0.714	0.066	0.033	0.044	No
rs675209	6	*RREB1/RREB1*	T	0.012	0.044	0.792	-0.024	0.044	0.593	No
rs742132	6	*LRRC16A/LRRC16A*	A	0.005	0.077	0.950	0.003	0.077	0.966	No

Supplementary Table 6 - 5: A summary of the pleiotropic effect of urate SNPs on lipids.

SNPs	Chr	Closest/GRAIL gene	Effect allele	TC (n=94,595)			LDL-c (n=89,888)			HDL-c (n=94,311)			ApoA (n=18,403)			ApoB (n=20,689)			Pleiotropy
				beta	se	p	beta	se	p	beta	se	p	beta	se	p	beta	se	p	
rs1260326	2	*GCKR/GCKR*	T	0.665	0.047	6.67E-46	0.268	0.048	2.58E-08	-0.147	0.045	1.24E-03	0.511	0.141	2.91E-04	0.867	0.135	1.31E-10	Yes
rs653178	12	*ATXN2/PTPN11*	T	-0.853	0.103	1.07E-16	-0.631	0.106	2.32E-09	-0.731	0.097	5.72E-14	-0.731	0.296	0.013	-0.162	0.287	0.573	Yes
rs17050272	2	*INHBB/INHBB*	A	-0.573	0.159	3.27E-04	-0.668	0.162	3.84E-05	0.027	0.151	0.858	-0.205	0.293	0.485	-0.367	0.279	0.188	Yes
rs642803	11	*OVOL1/LTBP3*	T	-0.281	0.084	7.76E-04	-0.270	0.086	0.002	-0.342	0.079	1.54E-05	-0.416	0.237	0.079	0.039	0.237	0.867	Yes
rs3741414	12	*INHBC/INHBE*	T	0.118	0.059	0.046	0.224	0.061	2.18E-04	-0.417	0.056	1.36E-13	0.010	0.172	0.953	0.430	0.172	0.012	Yes
rs1178977	7	*BAZ1B/MLXIPL*	A	0.196	0.090	0.029	-0.068	0.094	0.469	-0.632	0.088	6.88E-13	-0.336	0.279	0.229	0.650	0.265	0.014	Yes
rs6770152	3	*SFMBT1/MUSTN1*	T	-0.302	0.108	0.005	-0.244	0.110	0.027	-0.142	0.102	0.165	-0.143	0.226	0.527	-0.060	0.214	0.779	No
rs7224610	17	*HLF/HLF*	A	-0.287	0.137	0.036	-0.258	0.139	0.064	-0.105	0.129	0.414	0.305	0.281	0.279	0.591	0.267	0.027	No
rs729761	6	*VEGFA/VEGFA*	T	0.243	0.128	0.058	0.254	0.133	0.055	-0.317	0.122	0.009	-0.166	0.256	0.516	0.472	0.256	0.066	No
rs17786744	8	*STC1/STC1*	A	0.300	0.168	0.074	0.319	0.171	0.062	-0.213	0.158	0.178	0.470	0.344	0.172	0.319	0.333	0.338	No
rs1165151	6	*SLC17A1/SLC17A3*	T	-0.087	0.055	0.117	-0.055	0.057	0.327	-0.124	0.052	0.018	-0.144	0.116	0.213	0.012	0.112	0.916	No
rs2941484	8	*HNF4G/HNF4G*	T	-0.159	0.104	0.126	-0.080	0.108	0.462	-0.031	0.098	0.755	0.205	0.216	0.342	-0.157	0.205	0.444	No
rs11264341	1	*TRIM46/PKLR*	T	-0.154	0.113	0.171	-0.158	0.115	0.167	0.100	0.106	0.347	0.186	0.222	0.402	-0.096	0.211	0.649	No

SNPs	Chr	Closest/GRAIL gene	Effect allele	TC (n=94,595)			LDL-c (n=89,888)			HDL-c (n=94,311)			ApoA (n=18,403)			ApoB (n=20,689)			Pleiotropy
				beta	se	p	beta	se	p	beta	se	p	beta	se	p	beta	se	p	
rs1171614	10	*SLC16A9/SLC16A9*	T	0.120	0.093	0.197	0.077	0.096	0.422	0.020	0.089	0.820	0.179	0.184	0.332	-0.019	0.184	0.919	No
rs2231142	4	*ABCG2/ABCG2*	T	0.031	0.026	0.234	0.044	0.027	0.106	-0.056	0.025	0.027	-0.054	0.096	0.569	-0.024	0.087	0.786	No
rs2079742	17	*BCAS3/C17orf82*	T	0.163	0.149	0.275	0.086	0.153	0.573	0.214	0.141	0.130	0.008	0.273	0.977	-0.275	0.273	0.314	No
rs1394125	15	*UBE2Q2/NRG4*	A	0.135	0.137	0.326	0.135	0.140	0.334	-0.100	0.130	0.443	-0.280	0.279	0.316	-0.195	0.262	0.456	No
rs6598541	15	*IGF1R/IGF1R*	A	-0.070	0.084	0.402	-0.050	0.089	0.573	-0.248	0.082	0.002	-0.231	0.250	0.355	-0.009	0.239	0.969	No
rs478607	11	*NRXN2/SLC22A12*	A	0.110	0.146	0.449	0.150	0.150	0.317	-0.188	0.138	0.173	0.158	0.271	0.560	0.205	0.270	0.449	No
rs10821905	10	*A1CF/ASAH2*	A	0.085	0.130	0.514	0.051	0.132	0.700	0.019	0.121	0.876	-0.121	0.232	0.602	-0.237	0.231	0.305	No
rs2307394	2	*ORC4L/ACVR2A*	T	-0.094	0.151	0.534	-0.071	0.154	0.643	-0.083	0.143	0.562	0.530	0.320	0.098	-0.002	0.309	0.994	No
rs164009	17	*QRICH2/PRPSAP1*	A	0.097	0.183	0.597	-0.072	0.186	0.697	0.097	0.169	0.568	-0.204	0.379	0.591	0.467	0.361	0.195	No
rs7193778	16	*NFAT5/NFAT5*	T	0.081	0.157	0.608	-0.028	0.160	0.862	-0.253	0.147	0.085	-0.499	0.321	0.121	0.642	0.309	0.038	No
rs7188445	16	*MAF/MAF*	A	-0.078	0.175	0.655	-0.019	0.178	0.916	0.091	0.163	0.577	-0.445	0.355	0.209	-0.851	0.339	0.012	No
rs675209	6	*RREB1/RREB1*	T	0.035	0.092	0.704	-0.083	0.094	0.378	0.056	0.087	0.525	0.240	0.181	0.185	-0.113	0.173	0.513	No
rs12498742	4	*SLC2A9/SLC2A9*	A	0.005	0.016	0.747	0.008	0.016	0.617	-0.019	0.015	0.186	-0.027	0.033	0.425	0.033	0.033	0.320	No
rs10480300	7	*PRKAG2/PRKAG2*	T	0.053	0.178	0.766	0.072	0.184	0.697	-0.022	0.169	0.897	0.174	0.409	0.671	-0.277	0.388	0.475	No

SNPs	Chr	Closest/GRAIL gene	Effect allele	TC (n=94,595)			LDL-c (n=89,888)			HDL-c (n=94,311)			ApoA (n=18,403)			ApoB (n=20,689)			Pleiotropy
				beta	se	p	beta	se	p	beta	se	p	beta	se	p	beta	se	p	
rs2078267	11	*SLC22A11/SLC22A11*	T	-0.013	0.067	0.848	-0.054	0.068	0.428	0.074	0.062	0.227	-0.012	0.131	0.927	0.197	0.131	0.133	No
rs17632159	5	*TMEM171/TMEM171*	C	-0.026	0.155	0.865	-0.113	0.161	0.481	0.342	0.145	0.018	0.224	0.319	0.482	0.246	0.309	0.427	No
rs1471633	1	*PDZK1/PDZK1*	A	-0.008	0.089	0.926	0.054	0.090	0.549	-0.190	0.082	0.020	-0.147	0.172	0.392	0.306	0.163	0.061	No
rs742132	6	*LRRC16A/LRRC16A*	A	-0.011	0.163	0.944	-0.049	0.166	0.769	-0.009	0.151	0.955	-0.251	0.337	0.457	-0.360	0.320	0.260	No

Supplementary Table 6 - 6: A summary of the pleiotropic effect of urate SNPs on glucose.

SNP	Chr	Closest/GRAIL gene	Effect allele	Fasting glucose (n=15,234)			2hr glucose (n=58,074)			Glycoproteins (n=18,732)			Pleiotropy
				beta	se	p	beta	se	p	beta	se	p	
rs1260326	2	*GCKR/GCKR*	T	1.182	0.247	1.67E-06	-0.416	0.040	5.57E-25	-0.078	0.151	0.603	Yes
rs6598541	15	*IGF1R/IGF1R*	A	-0.682	0.455	0.134	0.273	0.075	2.77E-04	0.297	0.251	0.236	Yes
rs1171614	10	*SLC16A9/SLC16A9*	T	-0.297	0.392	0.448	-0.085	0.057	0.134	0.747	0.192	1.01E-04	Yes
rs1394125	15	*UBE2Q2/NRG4*	A	-1.302	0.512	0.011	0.009	0.084	0.912	-0.302	0.295	0.306	No
rs17050272	2	*INHBB/INHBB*	A	1.568	0.622	0.012	0.135	0.097	0.165	0.179	0.298	0.549	No
rs7224610	17	*HLF/HLF*	A	1.158	0.500	0.021	0.018	0.084	0.827	0.532	0.298	0.074	No
rs478607	11	*NRXN2/SLC22A12*	A	-1.083	0.521	0.038	0.025	0.088	0.775	0.189	0.276	0.493	No
rs653178	12	*ATXN2/PTPN11*	T	0.833	0.528	0.114	-0.053	0.089	0.553	-0.044	0.302	0.883	No
rs3741414	12	*INHBC/INHBE*	T	-0.479	0.324	0.139	0.014	0.054	0.792	0.083	0.177	0.638	No
rs6770152	3	*SFMBT1/MUSTN1*	T	0.563	0.396	0.155	0.044	0.065	0.498	0.490	0.226	0.030	No
rs7188445	16	*MAF/MAF*	A	-0.813	0.625	0.194	-0.191	0.103	0.065	-0.601	0.355	0.091	No
rs10480300	7	*PRKAG2/PRKAG2*	T	-0.719	0.656	0.273	-0.313	0.109	0.004	0.495	0.409	0.226	No
rs2231142	4	*ABCG2/ABCG2*	T	-0.150	0.145	0.302	0.031	0.024	0.199	0.104	0.093	0.264	No
rs1165151	6	*SLC17A1/SLC17A3*	T	0.207	0.207	0.317	0.068	0.034	0.042	0.112	0.116	0.335	No
rs164009	17	*QRICH2/PRPSAP1*	A	0.655	0.655	0.317	0.086	0.110	0.435	-0.040	0.380	0.915	No
rs7193778	16	*NFAT5/NFAT5*	T	0.447	0.574	0.437	0.136	0.096	0.155	0.050	0.324	0.878	No
rs1178977	7	*BAZ1B/MLXIPL*	A	-0.340	0.480	0.479	-0.122	0.080	0.127	-0.268	0.279	0.338	No
rs2307394	2	*ORC4L/ACVR2A*	T	-0.400	0.571	0.484	-0.191	0.094	0.042	-0.252	0.339	0.457	No
rs11264341	1	*TRIM46/PKLR*	T	-0.292	0.417	0.484	-0.075	0.069	0.275	-0.355	0.222	0.109	No
rs2078267	11	*SLC22A11/SLC22A11*	T	-0.154	0.244	0.528	-0.067	0.041	0.104	-0.027	0.137	0.843	No
rs1471633	1	*PDZK1/PDZK1*	A	-0.197	0.311	0.528	0.036	0.051	0.478	-0.035	0.173	0.839	No
rs742132	6	*LRRC16A/LRRC16A*	A	-0.371	0.600	0.536	0.089	0.097	0.362	-0.219	0.341	0.521	No
rs2941484	8	*HNF4G/HNF4G*	T	0.180	0.388	0.643	0.067	0.063	0.287	-0.021	0.219	0.924	No
rs675209	6	*RREB1/RREB1*	T	-0.156	0.349	0.656	0.130	0.057	0.023	0.215	0.183	0.240	No
rs2079742	17	*BCAS3/C17orf82*	T	0.235	0.529	0.657	-0.059	0.090	0.514	-0.622	0.274	0.023	No
rs17786744	8	*STC1/STC1*	A	0.255	0.613	0.678	0.132	0.100	0.186	-0.187	0.344	0.588	No
rs729761	6	*VEGFA/VEGFA*	T	-0.150	0.478	0.754	-0.202	0.078	0.010	-0.347	0.257	0.177	No

SNP	Chr	Closest/GRAIL gene	Effect allele	Fasting glucose (n=15,234)			2hr glucose (n=58,074)			Glycoproteins (n=18,732)			Pleiotropy
				beta	se	p	beta	se	p	beta	se	p	
rs17632159	5	*TMEM171/TMEM171*	C	-0.161	0.579	0.782	-0.050	0.092	0.587	-0.093	0.317	0.769	No
rs10821905	10	*A1CF/ASAH2*	A	0.045	0.453	0.920	0.079	0.075	0.294	-0.413	0.241	0.087	No
rs642803	11	*OVOL1/LTBP3*	T	-0.014	0.442	0.975	0.014	0.072	0.847	0.068	0.249	0.785	No
rs12498742	4	*SLC2A9/SLC2A9*	A	-0.001	0.058	0.985	0.000	0.009	0.978	0.029	0.033	0.376	No

Supplementary Table 6 - 7: Phenotypes associated with the weighted GRS of urate in TreeWAS analysis (PP>0.95).

ICD-10 coding	Disease description	max_b[†]	b_ci_lhs[†]	b_ci_rhs[†]	OR (95%CI)	PP[*]
M10	M10 Gout	1.640	1.515	1.765	5.16 (4.55, 5.84)	1.000
M100	M10.0 Idiopathic gout	1.640	1.515	1.765	5.16 (4.55, 5.84)	0.993
M1007	M10.07 Idiopathic gout (Ankle and foot)	1.640	1.515	1.765	5.16 (4.55, 5.84)	0.993
M1099	M10.99 Gout, unspecified (Site unspecified)	1.640	1.515	1.765	5.16 (4.55, 5.84)	1.000
M109	M10.9 Gout, unspecified	1.640	1.515	1.765	5.16 (4.55, 5.84)	1.000
M1097	M10.97 Gout, unspecified (Ankle and foot)	1.640	1.515	1.765	5.16 (4.55, 5.84)	1.000
M1096	M10.96 Gout, unspecified (Lower leg)	1.640	1.515	1.765	5.16 (4.55, 5.84)	1.000
M1094	M10.94 Gout, unspecified (Hand)	1.640	1.515	1.765	5.16 (4.55, 5.84)	0.993
M1090	M10.90 Gout, unspecified (Multiple sites)	1.640	1.515	1.765	5.16 (4.55, 5.84)	0.985
M109_int	M10.9 Gout, unspecified_int	1.640	1.515	1.765	5.16 (4.55, 5.84)	1.000
M100_int	M10.01 Idiopathic gout_int	1.640	1.515	1.765	5.16 (4.55, 5.84)	1.000
Chapter IX	Chapter IX Diseases of the circulatory system	0.070	0.055	0.085	1.07 (1.06, 1.09)	1.000
Block I10-I15	I10-I15 Hypertensive diseases	0.070	0.055	0.085	1.07 (1.06, 1.09)	1.000
I10	I10 Essential (primary) hypertension	0.070	0.055	0.085	1.07 (1.06, 1.09)	1.000
Block I20-I25	I20-I25 Ischaemic heart diseases	0.070	0.055	0.085	1.07 (1.06, 1.09)	1.000
I20	I20 Angina pectoris	0.070	0.055	0.085	1.07 (1.06, 1.09)	0.994
I209	I20.9 Angina pectoris, unspecified	0.070	0.055	0.085	1.07 (1.06, 1.09)	0.972
I21	I21 Acute myocardial infarction	0.070	0.055	0.085	1.07 (1.06, 1.09)	0.994
I219	I21.9 Acute myocardial infarction, unspecified	0.070	0.055	0.085	1.07 (1.06, 1.09)	0.966
I25	I25 Chronic ischaemic heart disease	0.070	0.055	0.085	1.07 (1.06, 1.09)	1.000
I251	I25.1 Atherosclerotic heart disease	0.070	0.055	0.085	1.07 (1.06, 1.09)	0.999
I252	I25.2 Old myocardial infarction	0.070	0.055	0.085	1.07 (1.06, 1.09)	0.987
Block I30-I52	I30-I52 Other forms of heart disease	0.070	0.055	0.085	1.07 (1.06, 1.09)	0.999
I50	I50 Heart failure	0.070	0.055	0.085	1.07 (1.06, 1.09)	0.994

ICD-10 coding	Disease description	max_b†	b_ci_lhs†	b_ci_rhs†	OR (95%CI)	PP*
I501	I50.1 Left ventricular failure	0.070	0.055	0.085	1.07 (1.06, 1.09)	0.966
Block I60-I69	I60-I69 Cerebrovascular diseases	0.070	0.055	0.085	1.07 (1.06, 1.09)	0.991
I63	I63 Cerebral infarction	0.070	0.055	0.085	1.07 (1.06, 1.09)	0.951

* PP, posterior probability for the beta (β) estimate in the tree analysis not being zero.
† max_b: maximum a posteriori effect estimate (beta) and the 95% credible interval (max_b_lhs, max_b_rhs).

Supplementary Table 6 - 8: Sensitivity analysis by using pleiotropic loci on metabolic traits.

Disease outcomes	GRS of all-urate loci (n=31)			GRS of urate-specific loci (n=14)			GRS of urate-obesity pleiotropic loci (n=10)			GRS of urate-BP pleiotropic loci (n=10)			GRS of urate-lipid pleiotropic loci (GRS=6)			GRS of urate-glucose pleiotropic loci (GRS=3)		
	OR (95%CI)	*p*	*FDR*	*OR (95%CI)*	*p*	*FDR*	*95%CI*	*p*	*FDR*	*95%CI*	*p*	*FDR*	*95%CI*	*p*	*FDR*	*95%CI*	*p*	*FDR*
Gout	5.37 (4.67, 6.18)	4.27E-123	TRUE	3.77 (3.19, 4.46)	4.42E-54	TRUE	12.82 (9.91, 16.59)	5.10E-84	TRUE	5.15 (3.26, 8.15)	2.32E-12	TRUE	9.52 (6.05, 14.99)	2.19E-22	TRUE	10.83 (6.37, 18.41)	1.38E-18	TRUE
Inflammatory polyarthropathies	1.27 (1.21, 1.34)	4.97E-19	TRUE	1.22 (1.15, 1.30)	6.45E-10	TRUE	1.57 (1.40, 1.76)	3.39E-14	TRUE	1.52 (1.26, 1.84)	1.27E-05	TRUE	1.52 (1.26, 1.83)	1.25E-05	TRUE	1.33 (1.07, 1.66)	0.010	FALSE
Hypertension	1.07 (1.05, 1.11)	6.02E-07	TRUE	1.03 (1.00, 1.07)	0.075	FALSE	1.14 (1.06, 1.22)	1.42E-04	TRUE	1.72 (1.55, 1.92)	2.13E-23	TRUE	1.45 (1.31, 1.61)	3.99E-12	TRUE	1.10 (0.97, 1.24)	0.138	FALSE
Essential hypertension	1.08 (1.05, 1.11)	6.26E-07	TRUE	1.03 (1.00, 1.07)	0.074	FALSE	1.14 (1.07, 1.22)	1.37E-04	TRUE	1.72 (1.55, 1.91)	2.87E-23	TRUE	1.45 (1.31, 1.61)	4.08E-12	TRUE	1.10 (0.97, 1.24)	0.146	FALSE
Coronary atherosclerosis	1.10 (1.05, 1.14)	1.17E-05	TRUE	1.05 (1.00, 1.11)	0.052	FALSE	1.18 (1.07, 1.30)	5.96E-04	FALSE	1.38 (1.18, 1.61)	3.37E-05	TRUE	1.80 (1.55, 2.09)	1.35E-14	TRUE	1.45 (1.21, 1.72)	3.89E-05	TRUE
Gouty arthropathy	5.10 (2.45, 10.66)	1.39E-05	TRUE	4.38 (1.77, 10.82)	0.001	FALSE	9.83 (2.50, 38.67)	1.08E-03	FALSE	1.94 (0.17, 21.86)	0.592	FALSE	3.28 (0.30, 36.00)	0.331	FALSE	12.57 (0.75, 209.57)	0.078	FALSE
Chronic ischaemic heart disease	1.09 (1.05, 1.14)	1.52E-05	TRUE	1.05 (1.00, 1.10)	0.057	FALSE	1.18 (1.07, 1.30)	5.79E-04	FALSE	1.37 (1.18, 1.59)	5.49E-05	TRUE	1.79 (1.54, 2.09)	2.64E-14	TRUE	1.44 (1.20, 1.71)	5.81E-05	TRUE
Ischaemic Heart Disease	1.09 (1.05, 1.14)	1.73E-05	TRUE	1.05 (1.00, 1.10)	0.060	FALSE	1.18 (1.07, 1.30)	6.64E-04	FALSE	1.37 (1.17, 1.59)	5.61E-05	TRUE	1.79 (1.54, 2.08)	3.61E-14	TRUE	1.43 (1.20, 1.71)	6.58E-05	TRUE
Myocardial infarction	1.14 (1.07, 1.22)	5.23E-05	TRUE	1.05 (0.97, 1.14)	0.205	FALSE	1.30 (1.12, 1.50)	4.54E-04	FALSE	1.65 (1.30, 2.09)	3.41E-05	TRUE	2.31 (1.83, 2.91)	2.27E-12	TRUE	1.75 (1.33, 2.30)	6.10E-05	TRUE
Pyogenic arthritis	2.10 (1.41, 3.13)	2.87E-04	TRUE	1.73 (1.07, 2.78)	0.024	FALSE	2.66 (1.17, 6.08)	2.00E-02	FALSE	1.71 (0.43, 6.80)	0.444	FALSE	9.58 (2.44, 37.68)	0.001	FALSE	6.13 (1.24, 30.22)	0.026	FALSE
Circulatory disease	1.04 (1.02, 1.07)	3.29E-04	TRUE	1.02 (0.99, 1.05)	0.258	FALSE	1.10 (1.04, 1.16)	6.94E-04	FALSE	1.31 (1.20, 1.43)	1.59E-09	TRUE	1.29 (1.18, 1.41)	9.47E-09	TRUE	1.17 (1.05, 1.29)	0.003	FALSE
Disorders of metabolism	1.07 (1.03, 1.11)	3.33E-04	TRUE	1.03 (0.99, 1.08)	0.157	FALSE	1.17 (1.08, 1.27)	1.06E-04	TRUE	1.12 (0.98, 1.27)	0.100	FALSE	1.58 (1.39, 1.80)	3.09E-12	TRUE	1.52 (1.30, 1.76)	6.35E-08	TRUE
Hypercholesterolaemia	1.08 (1.04, 1.12)	3.34E-04	TRUE	1.00 (0.95, 1.05)	0.913	FALSE	1.32 (1.20, 1.45)	3.53E-09	TRUE	1.23 (1.06, 1.43)	0.006	FALSE	1.90 (1.64, 2.20)	8.73E-18	TRUE	1.84 (1.55, 2.19)	3.01E-12	TRUE

GRS, genetic risk score; FDR, false discovery rate; OR, odds ratio.

7 DISCUSSION

7.1 Introduction

In this thesis, background information regarding uric acid metabolism, hyperuricaemia, and the genetic determinants of SUA level have been presented in ***Chapter 1***; the aims and objectives of this thesis have been outlined in ***Chapter 2***; a systematic literature review on the range of health outcomes related to SUA level has been summarised in ***Chapter 3***; the characteristics of UK biobank cohort and the manipulation process of UK Biobank data have been described in ***Chapter 4***; ***Chapter 5*** presents a MR-PheWAS analysis using the interim release of UK Biobank data and describes the methods and results, and interprets the findings; similarly, ***Chapter 6*** presents an enlarged PWMR analysis using the full UK Biobank cohort and describes the methods and results, and interprets the findings.

In this chapter, I will firstly discuss the methodological issues that have not been fully elaborated in the earlier chapters. I will also present and discuss the comments that have been proposed by the co-authors of the publications or the peer reviewers of the journals that this work has been submitted to. I will finally draw conclusions based on the findings of this thesis and provide suggestions on future research focus.

7.2 Methodological and analytical issues

The first part of this chapter discusses the main issues of the applied methodologies. These include (i) umbrella review; (ii) PheWAS approach: the PheCODE schema; (iii) MR methods: Two-stage MR, Wald Ratio MR, MR IVW and MR Egger; (iv) TreeWAS: the Bayesian analysis framework. Then, the common analytical issues of these methods are presented and the corresponding strategies for dealing with these issues are explained and discussed. These include (i) study population: why selecting the unrelated white British population for analysis instead of a broader target population of European descent; (ii) genetic instruments: why using the multiple genetic instruments rather than the single variant within the gene of which the function is well understood; (iii) covariates: why adjusting the PheWAS analysis only for the PCs and BMI but not for other potential confounding factors (e.g., sex, age, renal function); (iv) case ascertainment: how to define the individual as being a case or a control.

7.2.1 Umbrella review

To provide the research community with a comprehensive overview of the entirety of the published literature in relation to serum urate levels, I performed an umbrella review to

comprehensively assess the evidence available on the association between serum urate across all reported outcomes based on systematic reviews/meta-analyses of observational studies and RCTs as well as Mendelian randomisation (MR) studies. The main methodological issues have been fully discussed in *Chapter 3, Section 3.5 "Discussion"*, hereby, I only discuss a few remaining issues.

7.2.1.1 Selection of eligible studies

The methodology of umbrella review generally builds on the systematic analysis of meta-analyses. The selection of studies for inclusion in this umbrella review is relatively asymmetrical with the inclusion of systematic reviews and meta-analyses of observational studies, meta-analyses of RCTs, and finally, individual MR studies. The reason I sought to collect information from systematic reviews (not only meta-analyses) of observational studies was to map the breadth of outcomes investigated in relation to serum urate levels. Even though the systematic reviews were included in the umbrella review, they were not taken forward for quantitative analysis. For MR studies, only one meta-analysis was available, thus I decided to include individual studies instead. I sought to collect information from only systematic meta-analyses instead of all meta-analyses, this was because I aimed to keep the umbrella review to be systematic, objective and transparent. For meta-analyses which are not strictly systematic, there is always an unavoidable layer of subjectivity in making decisions about which papers should be included, and how the results should be interpreted and/or discussed. Nevertheless, there are concerns that some non-systematic meta-analyses cover questions that have not been addressed by systematic ones; if not including them, the breadth of health outcomes is limited. This has been discussed in *Chapter 3, Section "3.5 Discussion"* as a limitation of this umbrella review.

Additionally, another issue about the selection of the largest/latest meta-analysis when multiple (overlapping) meta-analyses exist for the same population/outcome has been questioned. This approach doesn't assume that the larger the meta-analyses, the better the quality, although there is some empirical evidence that this assumption is adequate within RCTs. Due to the large number of meta-analyses identified, it would be challenging to perform formal quantitative appraisal for all included meta-analyses and their component studies. Instead, I chose to present the main results of all identified meta-analyses and compare the findings for outcomes with more than one meta-analysis, and select one meta-analysis for each outcome to do the credibility assessment. The reason I selected the largest/latest meta-analysis for overlapping outcomes was because I expected that individual studies included in multiple meta-analyses should be considerably overlapped (if they

addressed the same question in the same population) and any difference should be due to new component studies being published and then being included in the most recent meta-analysis. However, there were some exceptions when the latest MA was not the largest and this was when the latest meta-analyses only included prospective studies whereas the largest meta-analyses included both prospective and retrospective studies. In this case I always selected the meta-analyses including the largest number of prospective studies. I then compared the consistency of the findings among these overlapping meta-analyses, aiming to see if the significance and/or direction of the association changed when new individual studies are added.

7.2.1.2 Criteria for credibility assessment

There were some criticisms that the criteria used for credibility assessment was largely based on statistical concepts (p-values, 95%PI, heterogeneity, *etc*.), highly dependent on sample size and the number of included studies. I realised that the credibility assessment based on statistical concepts is not sufficient to capture all features of the quality of meta-analyses, but these quantitative metrics are valid in identifying major biases. Considering that this umbrella review included 175 meta-analyses and each meta-analysis included multiple individual studies, it would be challenging to additionally perform a formal quality appraisal for all included meta-analyses and their component studies. It has been acknowledged that the assessment of the quality of the individual studies was beyond the scope of this umbrella review and it should be the aim of the original meta-analyses. Therefore, I examined the grading quality reported in the original meta-analyses (if any), but found quite a few meta-analyses using the GRADE (or other equivalent system) to assess the quality of their individual studies. The lack of the quality assessment of the component studies was highlighted as a common issue for the published meta-analyses. To improve this methodological limitation, I therefore applied a two-step assessment: I firstly performed quantitative assessment for all meta-analyses, and the statistical metrics filtered out meta-analyses with obvious/serious biases (Class III and IV); I then re-assessed the remaining meta-analyses (with high level evidence - Class I and II) on the basis of study design (case-control, retrospective, or prospective) and also by assessing the eligibility/quality of individual studies; and then conducted further evaluation on meta-analyses with high ranking evidence (Class I and II) to capture other quality issues that could have been missed in the quantitative assessment. In principle, although there are limitations, the methodologies I applied for credibility assessment can identify major biases and capture the essential features

of the quality of the included meta-analyses. Further evaluation for the meta-analyses which already have obvious/serious biases, would not add much value to the assessment.

7.2.1.3 Interpreting different types of evidence

In this umbrella review, different types of evidence (i.e., observational studies, RCTs and Mendelian randomisation studies) were incorporated. Although none of these study types are infallible, all are able to provide useful information about causal inference and can complement each other to achieve increasing certainty about causality. Observational studies examined the association between the exposure and the outcome and tested whether the association is caused by chance, bias, or confounding, but they are typically affected by residual confounding, undetected bias, or reverse causality, which may generate associations that are not reliable indicators of causality. RCT aimed to obtain evidence of a causal effect of a treatment or intervention on a disease process by eliminating many of possible biases and confounding factors, but they are limited by the issues of non-adherence to the assigned intervention, limited external validity, short term intervention effects, and non-retention, which can all render the results invalid or questionable. MR studies provided a cost-effective analogy to a RCT by using genetic variants as proxies to test the causality of an association between exposure and outcome; they are generally not influenced by the common confounding factors and not seriously affected by reverse causality, but do rely on several assumptions (the genetic instruments should be associated with the exposure of interest, they should not be associated with known confounders, and they should affect the outcome solely through the exposure) that can be hard to identify and control, and lack power when the proportion of trait variance explained by the genetic instruments is small. It is recognised that different types of epidemiological studies have specific strengths and weaknesses that can be seen as complementary, this study would therefore benefit from clarification as to the type of questions that each type of study is likely to address in terms of potential bias, generalisability and power limitation. Given the differences between different designs, I thus assessed each type of evidence separately and made specific conclusions for each study design type (the overall conclusion was a simple summary of specific conclusions). Specific conclusions were drawn for each of the three study types. In particular: 1) no association from meta-analyses of observational studies was classified as convincing; 2) only one outcome from meta-analyses of RCTs (decreased risk of nephrolithiasis recurrence with SUA-lowering treatment) had $p<10^{-3}$, 95% prediction interval excluding the null and no large heterogeneity or bias and was classified as convincing; 3) only one outcome from MR studies (increased risk of gout with high SUA levels) had $p<0.01$ with adequate power; the

overall conclusion of the chapter was a summary of specific conclusions: convincing evidence exists for only gout and nephrolithiasis. The comparison of the concordance across different study types aimed to see if evidence from RCTs and MR can strengthen the observed associations.

7.2.1.4 Strengths and limitations of umbrella review

The strengths and limitations of umbrella review have been fully discussed in *Chapter 3, Section "3.5 Discussion"*. In summary, strengths mainly includes:

(i) Providing an overview of the entirety of the published literature in relation to serum urate levels (reporting an impressive number [136 in total] of outcomes), and presented all published findings in a comprehensive way.

(ii) Collecting and evaluating evidence from multiple resources systematically, with recognising that different types of epidemiological studies were designed to address different questions but could be seen as complementary evidence for the same research topic.

(iii) Employing evidence classification criteria to identify biases and to capture the essential features of the quality of the published meta-analyses, to assess the credibility of the reported evidence.

(iv) Helping investigators to judge the relative priority of health outcomes related to serum urate levels and direct research or therapeutic efforts away from less important health outcomes in the future research and clinical management of diseases.

Limitations includes:

(i) Some meta-analyses were excluded from heterogeneity and bias tests because they did not provide adequate data to do the respective analyses.

(ii) Both asymmetry and excess statistical significance tests offer suggestions of bias, and not definitive proof thereof.

(iii) Effect inflation might affect even the results of the largest studies because often these studies were not necessarily very large or might have had inherent biases themselves. Thus, our estimates of the extent of excess statistical significance are probably conservative.

(iv) The quality of the individual component primary studies were not appraised because this was beyond the scope of this umbrella review. This was the aim of the original systematic reviews and meta-analyses, which should include an assessment of study quality and whether the study should be included in the quantitative calculations.

7.2.2 PheWAS approach

To complement GWAS, the concept of PheWAS has been developed. In essence, PheWAS is a reverse approach of GWAS, which provides a systematic method to explore the range of diseases and traits that are associated with a number of given SNPs or biomarkers (542). The concept of PheWAS is to perform a series of case-control tests on varying phenotypes to discover significant associations with pre-defined genetic variants or biomarkers of interest. Identifying the correlations among comprehensively collected phenotypes is believed to provide important information on the networks underlying human disease and health. PheWAS could be performed in a variety of datasets derived from electronic medical records (EMRs) or large population-based epidemiological studies (542). In particular, since the emergence of large biobanks (e.g., UK Biobank), in which a large volume of genotypic data is linked to extensive EMRs, it provides a unique opportunity to perform powerful PheWAS, even though there are some challenges to be overcome. The purpose of this section is to discuss the main technical issues of the PheWAS methodology and report the progress of PheWAS method that has been made based on the work of this thesis.

7.2.2.1 Phenotyping: ICD codes

The focus on a wider spectrum of phenotypes is an important aspect of the PheWAS design. Therefore, defining the individuals' phenotypes, termed as phenotyping, is the most critical step. Phenotyping has been defined as the precise and comprehensive analysis of phenotypic characteristics of an individual (543). The term of "phenome" has been used as a systematic and comprehensive set of phenotypes, including the clinical, biochemical and imaging traits. Unlike the genome, in which genotyping could be performed by reliable biological techniques, the measurement of phenome relies highly on the availability of health data. The most commonly used data fields for phenotyping are from the electronic medical records (EMRs) and the epidemiological survey data, which could be measured as either outcome phenotypes (binary disease status) or intermediate phenotypes (quantitative clinical variables) (542).

- **EMRs based phenotyping**

For EMRs based phenotyping, the most effective and straightforward way is to use the ICD codes. The ICD coding system is a medical classification list developed by the World Health Organisation (544). This system is designed to map health conditions to corresponding generic categories together with specific variations. The major categories are designed to include a set of similar diseases. The ICD has been applied as a health care classification

system, which provides a system of diagnostic codes for classifying diseases and a wide variety of signs, symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or disease (544).

The ICD is revised periodically and is currently in its 10th version (ICD-10) (545). Establishment of ICD-10 began in 1983 and completed in 1992. The ICD-10 was first mandated for use in the UK in 1995 (before 1995, the adopted version was ICD-9). In 2010, the UK Government made a commitment to update the UK version of ICD-10 every three years (546). From April 1st 2016, the ICD-10 5th Edition was applied as the mandated diagnostic classification within the UK (546). The current version of ICD-10 includes more than 14,400 different codes (permitting the tracking of many new diagnoses compared to ICD-9) and through the use of optional sub-classifications, the number of codes can be expanded to over 16,000 (544). The unchanged international version of ICD-10 is used in more than 110 countries, with a few exceptions. For example, the United States, Australia and Canada, have developed their own adaptations of ICD, which include more procedure codes for classification of operative or diagnostic procedures. The United States expanded the code set to a modified version, named ICD-10 Clinical Modification (ICD-10-CM), which has 70,000 different codes (547). Adoption of ICD-10-CM was slow in the United States and it has been used since October 1, 2015. The ICD-10-CM was provided by the National Centre for Health Statistics (NCHS) and the Centres for Medicare and Medicaid Services (CMS) and the use of ICD-10-CM codes are now mandated for all inpatient medical reporting requirements in the United States (547).

As known, the ICD-codes are organised in a hierarchical structure that the sibling ICD codes representing different subgroups belong to a similar set of disease group coded by their parent ICD code (544). Based on this hierarchical structure, a holistic phenotyping method was proposed for PheWAS, in which individual ICD codes at varying levels of phenotypic granularity are adopted to define cases and controls. For example, M10.1 and M10.2 are firstly classified into two different case groups, but then are combined into the same M10 case group. The advantage of this method is that the phenotyping is performed without making any assumptions about the similarity of outcomes and the ICD codes are treated as representing independent disease outcomes. However, given the fact that the majority of ICD codes are designed to include a set of similar diseases, this phenotyping method would greatly reduce the study power and increase the burden of multiple test.

To address these issues, the curated phenotyping method has been adopted to aggregate the ICD codes into an appropriate version for phenotyping. The curated phenotyping is built on

the backbone of the first three digital ICD codes and further refined based on clinical and biological knowledge (548). In principle, the fourth and fifth digital ICD codes, or sometimes even the third, representing diseases sharing the common aetiology are combined into one phenotype group. On the other hand, for some diseases with distinctly different aetiologies, like type 1 and type 2 diabetes, even if they are under the same three digit ICD code, their individual codes are retained and allocated into different phenotype groups. In addition, ICD codes, using the terms of "others" or "not elsewhere classified" are removed. Through the curated phenotyping, the number of cases for each phenotype is increased and the number of phenotypes examined in PheWAS is reduced, therefore, the multiple test burden is lower and the power to detect associations is higher due to increased number of cases. This phenotyping method also has its disadvantages, in which biases may be introduced by assuming that some diseases share the same aetiology. Despite this, curated phenotyping is still one of the most widely adopted methods in the published PheWAS. Based on the principles of curated phenotyping, a PheCODE schema has been developed for the researchers who are interested in the EMRs based phenotyping (details are discussed in the *Section 7.2.2.2 "PheCODE schema"*).

- **Epidemiological survey-based phenotyping**

Another important source of phenotypic data is the epidemiological surveys, in which a large number of phenotypes are collected. Compared with ICD codes, phenotyping in epidemiological data is even more challenging. Considering the relatively small sample size and the limited coverage of human diseases investigated in a single epidemiological survey, combination of phenotypic data from multiple studies is always required (548). Additionally, since the diseases investigated in epidemiological studies are not as well-structured and standardised as ICD codes, it is difficult to harmonise all phenotypes, and thus phenotypes are always manually binned into phenotype classifications (441). Although this method can reuse the epidemiological data to perform PheWAS, it is not the best option to harmonise the phenotypes of individuals.

7.2.2.2 Phenome framework: PheCODE schema

As discussed above, the ICD codes remain the most common data source of phenotyping. To address the need for an aggregation method, the electronic Medical Records and Genomics (eMERGE) developed a schema called "PheCODE" to represent disease phenotypes documented in the EMR (433). The PheCODE schema was first introduced in 2010, containing 733 distinct phenotype codes (459). This schema was developed based on the ICD-9-CM codes by combining one or more related ICD-9-CM codes into distinct diseases

or traits (459). The eMERGE group are continuously refining the PheCODE schema with additional clinical experts helping with revisions of different domains, such as cardiology and oncology. The latest version of the PheCODE system involved 1,866 hierarchical phenotype codes (433).

Although the PheCODE schema is effective at replicating genotype-phenotype associations, as explained above, the phenotyping algorithm developed by the US eMERGE group was based on the ICD-9-CM and it meant that the phenotyping algorithm could not be used directly for the UK Biobank data, in which the EMRs are largely encoded by the ICD-10 version. In order to adopt the PheCODE schema to define the phenome, the US eMERGE group in Vanderbilt University Medical Center (VUMC) helped to create the mapping of ICD-10 codes to the current PheCODE system and evaluated the coverage performance of PheCODE for EMRs documented in the UK Biobank. Although multiple ways were applied to create the mapping to phecodes, about 26.8% ICD-10 codes used in the UK Biobank could not be mapped to any phecode. An examination on these codes showed that the majority were encounter/procedural codes or supplementary codes that are not descriptions of specific phenotypes or diseases. Of the ICD-10 codes that were successfully mapped to phecodes, about 8,947 ICD-10 codes were mapped to at least one phecode, and 256 (2.9%) codes were mapped to more than one phecodes. For example, the ICD-10 code B21.1 was mapped to two phecodes: 071.1 [HIV infection, symptomatic] and 202.2 [Non-Hodgkins lymphoma]. In addition, ICD-10 codes involving in personal or family history were also largely unmapped, and this revealed a potential shortcoming of the current PheCODE system (i.e., the missing element of family history) and demonstrated an area of which the system should be expanded and improved.

7.2.2.3 Other phenotyping resources

Apart from the eMERGE network, there are several other notable efforts focused on developing a more unambiguous definition of phenome, including the work from the Human Disease Ontology (DO) resource (549) and the Human Phenotype Ontology (HPO) project (550).

The Human DO is a standards-based ontology that focuses on representing common and rare disease concepts captured from biomedical resources and organised by disease aetiology (549). The latest revision includes 8803 classes (terms) (6419 non-obsolete, 2384 obsolete) and provides textual definitions for 32% of DO classes (terms) (549). Human DO is a disease-focused database by design and thus it includes only concepts of disease. More specifically, it integrates disease concepts from ICD-9, the National Cancer Institute (NCI)

Thesaurus (551), SNOMED-CT (552) and MeSH (https://www.nlm.nih.gov/mesh/MBrowser.html) extracted from the UMLS (553) based on the UMLS CUI for each disease term; it also includes disease terms extracted directly from Online Mendelian Inheritance in Man (OMIM) (554), and the Experimental Factor Ontology (EFO, http://www.ebi.ac.uk/efo/). However, the Human DO does not include progression (early, late, metastasis, stages) or manifestations (transient, acute, chronic) of disease as part of the disease definition; and it does not include compound disease terms (those describing the combination of two disease terms) such as glaucoma associated with pupillary block either. Instead, these diseases are represented by two distinct disease terms. Details about the Human DO resource are available at http://www.disease-ontology.org.

The Human Phenotype Ontology (HPO) project includes a wide range of phenotypic abnormalities described in human diseases (550). At the time of writing this thesis, The HPO provided a structured, comprehensive and well-defined set of 10,088 classes/terms describing human phenotypic abnormalities. About 65% (6,603) of the classes are described by a detailed textual definition created by clinical experts. In addition, a logical definition is developed for 46% of all HPO classes using terms from ontologies for anatomy, cell types, function, embryology, pathology and other domains. The HPO is organised as three independent sub-ontologies covering different categories: the mode of inheritance, the onset and clinical course and the largest category of phenotypic abnormalities. Details about the HPO database are available at http://www.human-phenotype-ontology.org.

Compared to the other phenotyping resource (e.g., the HPO and Human DO), the PheCODE schema remains the most direct way to use the diagnosis codes and enables the performance of PheWAS by leveraging EMRs for high-throughput analysis. Additionally, machine-learning approaches like trained logistic regression models or support vector machines are also proposed to be used in phenotyping (548). Principle component analysis is another possible method to transform a number of phenotypic variables into a smaller number of phenotypic groups (555). Efforts are also made to evaluate the accuracy of combining various structured and unstructured data sources (549, 550). Hopefully, in the near future, more advanced phenotyping methods will be developed with collaboration from multiple organisations to discover disease subclasses and provide the means to better capture, store, exchange, and analyse phenotypic data.

7.2.3 TreeWAS method

A limitation of the conventional PheWAS using PheCODE schema is testing a tree-structured phenotype spectrum with a general linear model by assuming independence, although they are actually correlated. This would result in over correction of p value when adjusting for multiple testing. This issue was addressed by the alternative Bayesian analysis of a tree-structured phenotypic model (TreeWAS).

In TreeWAS analysis, phenotypes are organised into a tree structure following the hierarchical structure of ICD10 codes to better capture the underlying biological process affecting the origin and progression of disease. The associations between weighted GRS and phenotypes were tested by the Bayesian network analysis at both terminal and internal nodes of the tree structure. The expected degree of correlation between genetic coefficients across each node in the tree was determined by the prior probabilities. The coefficient at a parent node can either be inherited by a child node with a probability, or can transition to a new uncorrelated value with a probability. The transition probabilities controlled the Markov process and the likelihoods over the genetic coefficients were calculated across all clinical phenotypes using a dynamic programming model and the forward and backward algorithms. The value of TreeWAS lies in enhancing power to identify groups of endpoints affected by exploiting the encoding of medical ontologies. With taking into account the correlations among clinical phenotypes, this new approach is shown to increase statistical power by up to 20%.

Additionally, TreeWAS using a Bayesian network analysis has an advantage in detection of non-linear associations that could be missed by conventional PheWAS. PheWAS analysis tested the association by running a logistic regression under a hypothesis of a linear model, thus any association that is non-linear would be less likely to be detected. The association between serum urate level and the risk of cerebrovascular diseases (e.g., stroke) is probably a case of non-linear association, which was only observed in TreeWAS but not in PheWAS. Previous studies assessing the association between serum urate level and cerebrovascular diseases (e.g., stroke) reported conflicting results, where some studies found a positive association whereas other studies reported a negative association. Some studies suggested that high SUA level was neuroprotective and associated with better outcomes after acute ischaemic stroke, while other studies reported that high SUA level was injurious and had a statistically significant association with the high risk of stroke incidence and mortality (556-558). Given the contradictory role of urate as both an antioxidant and a pro-oxidant, a compromise hypothesis of U-shaped relationship has been proposed that suggests low serum

urate level may increase the risk of cerebrovascular disease as a result of inadequate anti-oxidative level and high serum urate level may function more as a pro-oxidant to increase the risk of stroke and other cerebrovascular diseases (558). If the presence of U-shaped relationship is true, the PheWAS analysis, in which association was tested under a hypothesis of a linear model, would be thereby unable to detect this association, however TreeWAS, testing the association without assumption of any linear model, reported a significant association.

7.2.4 MR methods

To interpret the observed PheWAS associations, I applied MR to determine whether there is any causal effect of SUA level on the identified diseases. MR is a method that uses genetic variants as instrumental variables (IVs) to examine the causal effect of an exposure on a disease (559). MR relies on the natural, random assortment of genetic variants during meiosis, which yields a random distribution of genetic variants in a population. Specially, individuals are naturally assigned at birth to inherit or not inherit the genetic variants that affect an intermediate phenotype (e.g., genetic variants raising SUA level); individuals who carry the risk genetic variants and those who do not are then followed up for the development of an outcome of interest. Ideally, as genetic variants are typically not associated with the common confounders and the transmission of genetic information is usually unidirectional, differences in the disease outcome between those who carry the variant and those who do not, can be attributed to the difference in the risk factor. This section will focus on the methodological issues that are prominent in the MR method.

For the MR analyses of SUA levels conducted in this thesis, a list of crucial issues with regards to the MR approaches are carefully and sufficiently assessed and discussed, in particular the assumptions that underpin the MR design, the use of multiple instrumental variables, the choice of statistical methods for robust effect estimation, the statistical power to detect a moderate/small causal effect, and the interpretation of associations due to pleiotropy or genetic linkage disequilibrium.

7.2.4.1 MR assumptions

There are three main assumptions that underpin the MR method (as briefly mentioned in ***Chapter 4***). These are:

(i) The genotype is associated with the exposure. This assumption can be verified by statistical analysis. Reporting guidelines for MR analysis recommend the use of the partial F-statistic as a measure of the strength of the association between the IV and the

exposure (560).

The F statistic is defined as the ratio of the mean square of the model (the first-stage regression of exposure on the IV) to the mean square of the error. In the MR reporting guidelines, a threshold of F <10 is typically used to define a 'weak IV'. This criterion is based on the observation that the F value greater than 11 ensures the relative bias to be <10% regardless of the number of IVs used in MR analysis (560). However, in the MR analyses of SUA level there are no biomarker data available for the SUA levels of the UK Biobank participants. Therefore, I calculated the statistical power to report the strength of IVs, given that the F statistic encompasses information on the first-stage R^2 (the proportion of variance in SUA level that is explained by IVs), the sample size and the number of IVs, which are also indicators of the statistical power of MR analysis (257). Power >80% was regarded as sufficient to detect the corresponding causal effect, otherwise, a statement on inadequate power was made to acknowledge the lack of strength to detect the causal effect. Details about the power estimation are further discussed in the *Section* "*7.2.4.3 Statistical power of MR*").

(ii) The genotype is associated with the outcome through the exposure of interest only. This assumption is likely be violated when the genotype has multiple (pleiotropic) effects, or when a nearby genetic variant in linkage disequilibrium with the IV affects the outcome in other ways than through the exposure of interest.

One of the primary concerns on evaluating this assumption is whether the genetic variants used as IVs are likely to be pleiotropic. Pleiotropy can affect the interpretation of MR studies in multiple ways: (i) a reverse pleiotropic effect can counteract the effect of the biomarker on the disease, which will give a null finding even when there is a true causal effect between the biomarker and the disease; or alternately (ii) a positive pleiotropic effect can result in a positive association between the genetic variant and the disease that is likely to be mistakenly interpreted as a causal effect. To validate this assumption, I examined the presence of pleiotropy and balanced out pleiotropic effects by using a more advanced statistical technique, the MR Egger analysis that has recently been proposed to account for unbalanced pleiotropy. Details about MR Egger method are discussed in the *Section* "*7.2.4.2 MR methods*".

Another concern about this assumption is genetic linkage disequilibrium, given the fact that SNPs located closely on a chromosome are usually inherited together. The closer the distance on a chromosome, the higher the extent of linkage disequilibrium. For example, a genetic variant that affects the exposure level of A (e.g., SUA) may be in linkage

disequilibrium with a SNP affecting the exposure level of B (e.g., LDL). If the exposure B is causally related to a disease outcome, the MR findings using the genetic variant in linkage disequilibrium with the genetic variant affecting exposure B would be mistakenly interpreted as that the exposure A being responsible for the corresponding disease outcome. To avoid any potential misinterpretation, the ideal IVs should be SNPs that localise in genomic regions without proximity to the loci that might affect the association of the SNP and the disease. However this is not always possible given the complexity of human genetic architecture. To validate this assumption, I applied the HEIDI test to examine if any of the observed associations are due to linkage disequilibrium instead of causality. Details on the HEIDI test have been elaborated in *Chapter 5, Section* "*5.3 Method*", "*5.5 Discussion*" and "*5.6 Supplementary information*")

(iii) The genotype is independent of other factors, which affect the outcome (independence assumption). This assumption would be violated if subgroups in the study population have both different genotype frequencies and different distributions of the outcome (population stratification), or if there is an association between the genetic instruments and confounders.

Population stratification refers to the confounding of genotypic associations by factors due to differences in subgroup populations. If the study population is not homogenous, any disease outcome that is at higher prevalence in one of the subpopulations may be associated with all SNPs that are more frequent in this group. To address this issue, I performed the SUA MR analyses in a very homogenous population (white British subset) and adjusted using the genetic principle components as covariates. For assessing the second common issue related to this assumption, I tested whether the genetic instruments of SUA level were associated with other common confounders such as BMI, smoking, assessment centre and any related confounders were adjusted as covariates in the analysis (Details about adjustment for covariates are discussed in the *Section* "*7.2.5.3 Adjustment for covariates*").

7.2.4.2 MR methods

A number of statistical methods have been proposed for MR to obtain robust causal estimates and/or to examine the instrumental variable assumptions (561). These include the standard MR methods (i.e., two-stage method, Wald Ratio method) that are most frequently used when the individual-level data (i.e., data on exposure, outcome and IVs) are available from a single population (one-sample MR), and the more complicated methods (i.e., inverse-variance weighted [IVW] MR, Egger MR) that are developed to integrate data from multiple samples (two-sample MR) and multiple IVs. In this section, I mainly discuss the mostly

common used methods as mentioned above, but it should be noted that this is not an exhaustive list of methods that are available for MR.

- ***Two-stage method***

One of the most frequently used statistical methods for MR is two-stage analysis. This involves two stages of regression. The first stage is a linear regression with the instrument (genotype) as the independent variable and the exposure as the dependent variable, which is then used to obtain the exposure levels predicted by the instruments. The second stage is a regression with the outcome as the dependent variable and the genetically determined exposure level as the independent variable. With continuous outcomes under a linear model, the two-stage method is also known as two-stage least squares (2SLS). With binary outcomes, the second-stage (exposure–outcome) regression is a logistic regression model. The causal estimate is the second-stage regression coefficient that is explained as the change in the outcome caused by a unit change in the exposure and the estimator is expressed as a causal relative risk or odds ratio. However, there is a debate on the concern that estimates for binary outcomes from this method is inflated, as the uncertainty in the first-stage regression is not accounted for and the non-linear model does not guarantee that the residuals from the second-stage regression are uncorrelated with the instruments (561, 562). Despite this, the two-stage estimator with a logistic regression second-stage model still provides a valid test for the null hypothesis.

- ***Wald Ratio method***

The Wald Ratio method, also known as the ratio of coefficients method, is the simplest way to estimate the causal effect of the exposure (X) on the outcome (Y). The ratio method typically uses a single IV (Z). If the coefficient of the regression of the exposure on the IV is denoted as b_{zx} and the coefficient of the regression of the outcome on the IV is denoted as b_{zy}, then the causal estimate could be calculated by the formula: ratio method estimate= b_{zy}/b_{zx} (563). Intuitively, this estimate could be explained as follows. The change in the outcome for a unit increase in the exposure is equal to the change in the exposure for a unit increase in the IV that is scaled to the change in the exposure for a unit increase in the IV. With a single IV, the causal estimate from the ratio method is the same as that of the 2SLS method (561). With multiple IVs, the 2SLS estimate could be viewed as a weighted mean of the ratio estimates calculated by using one instrument at a time, where the weights are determined by the corresponding coefficients obtained from the first-stage regression (561).

The MR methods discussed above, particularly the two-stage method, require individual-level data on genetic instruments, exposure and outcome from a single population. However,

in the case of SUA MR analysis performed in this thesis ,where the SUA levels were not yet available in UK Biobank, I therefore performed the alternative two-sample MR (2SMR) by using data from two independent populations, where the coefficient of the regression of SUA level on the genetic variant was taken from the summary-level GWAS data of European decent that were provided by the GUGC consortium (151) and the coefficient of the regression of outcome on the genetic variant was estimated by using the individual-level data from UK Biobank.

- ***Inverse-variance weighted method***

The standard statistical method for performing two-sample MR is the inverse variance weighted (IVW) approach which is developed by combining the concepts of the ratio method and meta-analysis (474). For a single IV, the causal estimate for each IV could be calculated by the ratio method. With multiple IVs, the ratio estimates from each genetic variant can be averaged using an inverse-variance weighted formula that is taken from meta-analysis, where the IV-specific causal estimates are equivalent to the study-specific estimates, and the weights are the inverse-variance weights (474). The causal effect of the exposure on the outcome is estimated using a weighted linear regression where the residual standard error is set to one and the intercept is set to zero (474). This weighted regression model is equivalent to performing a fixed-effect meta-analysis (564). When using a fixed-effect model for combining the IV-specific causal estimates, no heterogeneity between the causal estimates of the individual genetic variants is expected (564). When substantial heterogeneity is present, the MR IVW analysis using a fixed-effect model is not recommended. Additionally, the MR IVW method also relies highly on the three fundamental assumptions listed above. As emphasised, if the three assumptions (i)–(iii) hold, then the causal estimate from IVW MR is robust and unbiased. However, assumptions (ii)–(iii) are likely to be violated in the MR study using multiple genetic instruments. To validate these assumptions, I thus applied the MR Egger method to assess whether the genetic variants have any pleiotropic effect (directional pleiotropy) and to provide an unbiased causal estimate under weaker assumptions — the InSIDE (Instrument Strength Independent of Direct Effect) assumption (469).

- ***MR Egger method***

The MR Egger method is developed to provide robustness against misspecification of the MR assumptions (469). MR Egger is performed by a simple modification to the weighted linear regression described above. Instead of setting the intercept term to be zero, the intercept is estimated as part of the regression. If the genetic instruments are not pleiotropic,

then the intercept term should tend to be zero and the MR Egger estimate should be equal to the IVW estimate. Otherwise, if the genetic instruments are pleiotropic but the direct effects of the genetic variants on the outcome are independent of the associations of the genetic variants with the exposure (known as the InSIDE assumption), then the MR Egger regression will return a null estimate for the intercept (469). Under the InSIDE assumption, the intercept can be interpreted as an estimate of the average direct effects of the genetic variants on the outcome (469). If the average direct effect (intercept) is zero (referred as 'balanced pleiotropy'), and the InSIDE assumption is satisfied, then the estimates of MR Egger should be robust. If the intercept term differs from zero (the average direct effect is not zero), then the InSIDE assumption is violated (referred as 'directional pleiotropy'), and the MR estimate is biased. Hence, testing the intercept from the MR Egger analysis provides an assessment of the validity of the IV assumptions. Although the MR Egger is more robust in dealing with pleiotropy, this method is not infallible, and a number of methodological issues (i.e. the precision of the estimate, the influence of outlying variants, and the violation of the InSIDE assumption) have been proposed as limitations and these should be noted when interpreting the results (481).

The precision of the MR Egger estimate not only depends on the proportion of variance in the exposure explained by the genetic variants (measured as R^2 statistic), but also depends on the variability of the strength of individual genetic instruments (481). A precise MR Egger estimate requires the consistency of the causal estimates across the genetic variants. The use of pleiotropic variants, where heterogeneity between the causal estimates is observed, would result in over-dispersion in the MR Egger regression, in which a random-effects model is preferred. Therefore, the standard error of the causal estimate from the MR Egger method (random-effects model) is typically larger than that from the MR IVW method (fixed-effect model) and accordingly the 95%CI of the causal estimate from the MR Egger method is also wider than that from the MR IVW, which would result in an imprecise estimate.

Besides, the MR Egger estimate is easily influenced by any outlying variant (481). If one genetic variant has a much stronger association with the exposure than others, then this variant would have a larger influence on the coefficients in the MR Egger regression. As illustrated by the example of hypertension mentioned in *Chapter 5 Section "5.5 Discussion"*, it is found that the outlying variant (rs12498742 in *SLC2A9/SLC2A9* gene) that had the strongest association with SUA level showed a negative effect on hypertension and reversed the sign of the overall putative causal effect. In such a case, it implies that the InSIDE assumption is likely to be violated. The test for directional pleiotropy indicates that the genetic variants are not all valid instruments, thus the negative MR Egger estimate is highly

dubious, as the causal estimates from each variant are all positive.

Given the respective limitations and strengths of the MR IVW and MR Egger methods discussed above, the next important issue is the interpretation of the discordant results from MR Egger and MR IVW. As discussed above, if the intercept parameter is close to zero, then the MR Egger estimate should be close to the IVW estimate, but the 95%CI of the causal effect estimated from MR Egger method is wider than that estimated from the IVW method due to the imprecision of MR Egger regression (481). In this case, the MR Egger analysis does not contradict evidence for a causal effect estimated from MR IVW analysis. On the contrary, if the intercept parameter is not zero, it implies the MR assumptions are violated and the pleiotropic effects from multiple instruments cannot even be balanced, thus neither the MR IVW estimate nor the MR Egger estimate are robust.

In conclusion, it should be emphasised that the main point of the MR analysis is the assessment of the IV assumptions or the use of valid instruments. If the IV assumptions do not hold, then inferences from any analysis method will be unreliable. Although there are some situations where a particular method is more suitable than others (e.g., MR Egger method is more robust for the use of multiple instruments with balanced pleiotropy), there is no single MR method that is universally best. Instead, it makes sense for using different IV methods (e.g., MR Egger analysis) as sensitivity analyses to assess if the estimate given by a particular choice of MR method (e.g., MR IVW analysis) is credible or not.

7.2.4.3 MR statistical power

Statistical power is the probability that the null hypothesis (H0: no association) can be rejected if a specific alternative hypothesis (H1: a true association of the biomarker with disease risk) is true. The statistical power ranges from 0 to 1, and as statistical power increases, the probability of a type 2 error (false negative) decreases. In the published MR studies, investigators typically evaluate the F-statistic and the R^2 from the first-stage regression (exposure-genetic variant) to directly reflect the power of MR analysis. Recently, Brion *et al* develops a non-centrality parameter (NCP)-based approach for calculating power of an MR study (254). The F-statistic and the R^2 are strong determinants of the statistical power to detect a causal effect. Other determinants of statistical power include the sample size, the prevalence of the outcome in the study population and the true effect size of the exposure on the outcome. A formula for power estimation has been provided as an online web tool for researchers to perform calculation (http://glimmer.rstudio.com/kn3in/ mRnd/).

As genetic variants typically explain a small/moderate proportion of the variance in biomarkers, the statistical power to detect an association between the variant and outcome in

MR can be relatively low and usually a large sample size is required to achieve adequate statistical power. Estimating statistical power is important as it can inform whether a null finding is representative of a true null causal association, or simply a lack of power to detect an effect size of clinical interest. Statistical power can be improved by the use of multiple genetic instruments to increase the proportion of variance of explained by the biomarker, or by increasing the sample size.

Ideally, a power calculation is typically used in estimating sufficient sample size to achieve adequate power, thus it is usually performed before the study is carried out. However, in this thesis using data from UK Biobank cohort, where the sample size (and the number of cases for specific disease) is fixed, statistical power was calculated by using the NCP-based approach to investigate the minimum effect size that is likely to be detected. As the MR analysis is incorporated within the PheWAS analysis to investigate a wide range of phenotypes, the numbers of cases for different diseases can be considerably different depending on their prevalence. Therefore, the power of MR analyses for different diseases varies significantly from 1.0 to 0.01. The statistical power of MR studies to detect the specific causal effects have been noted and discussed in *Chapter 5, Section "5.3 Method", "5.5 Discussion"*.

7.2.5 Analytical issues

This section mainly discusses the analytical issues proposed by the peer reviewers and/or the co-authors of the publications derived from this thesis.

7.2.5.1 Study of white British population

When submitting the MR-PheWAS paper to the journal of *Annals of the Rheumatic Diseases*, one of reviewers proposed that "the use of the self-reported British (confirmed by PCA) is totally fine but a broader study population of European descent could add around 20K participants to increase the power and the use of interim release data sets a further limitation on the power of this study."

I agree with the reviewer that the use of interim release of data for MR-PheWAS analysis limited the power of this study and have acknowledged the proposed point as a limitation in the paper. The use of self-reported British (confirmed by PCA) was to minimise the influence of population stratification. I agree with the reviewer that the use of British population reduced the sample size when compared with the use of European descent, however, given the fact that the genetic determines of SUA level are diverse in different ethnicities and that the excluded samples consist of more controls than cases, I decided to use a homogenous British population after weighing the gains (minimise the influence of population stratification) and losses (lose study samples). Additionally, it should be noted that even the British people are not definitely homogenous, further adjustments for principal components of ancestry should be considered to minimise the influence of the population stratification.

7.2.5.2 Use of multiple instrumental variables

Another reviewer commented that "Clearly, pleiotropy is an issue, with some of the serum-associated PheWAS variants used (e.g. *GCKR*) associated with multiple other phenotypes. Therefore I wonder why a simpler approach, using genetic variation within the well understood *SLC2A9 gene*, with no evidence for pleiotropy, is not used in studies."

I agree that the *SLC2A9* gene suggested by the reviewer is a good candidate instrument, as it has the strongest association with SUA level and its biological function is well characterised as a urate transporter. So far as I know, there is no other disease (except gout) that has been reported to be associated with this locus (without observed evidence for pleiotropy), although it should be noted that the *SLC2A9* gene is also a glucose/fructose transporter, which may leave the possibility for unobserved pleiotropy. However, the main reason for using multiple genetic variants in MR was to increase the strength of genetic instrument, as

the 31 SNPs all together can explain about 7% of SUA variance, while the *SLC2A9* gene only accounted for less than 2% (151). An MR analysis using a weak instrument would easily suffer from bias and would tend to underestimate the true uncertainty, leading to inflated type 1 error rate (561).

Furthermore, the use of multiple genetic instruments would strengthen the causal inference by noting the disease outcomes associated with at multiple SUA risk loci. This could be analogous to the RCTs using different SUA-lowering drugs, which work through different mechanisms and have different potential side-effects but lower SUA to the same degree. If the different SUA-lowering drugs produce the same reduction in the disease risk, then it is less likely that this effect is through agent-specific effect of the drugs, instead it points to SUA lowering as being the key. For example, gout is associated with multiple independent SUA risk loci, pointing to the same underlying SUA-outcome (gout) association; intuitively it is less likely that the association is caused by the reintroduced confounding (e.g. pleiotropy or LD), unless the reintroduced confounding acts in a way influencing two unlinked genetic variants.

In addition to causality, I was also interested in pleiotropy, which may reflect common aetiology or biological pathways shared by the affected diseases. Using multiple genetic instruments allowed me to investigate pleiotropic associations of interest. As acknowledged, the use of multiple genetic instruments increases the risk to break the MR assumptions due to pleiotropy. To address this problem without vitiating the statistical power or losing the chance to investigate pleiotropic association, I therefore applied the advanced MR Egger analysis to correct for the observed and unobserved pleiotropy that was potentially caused by using multiple genetic instruments.

7.2.5.3 Adjustment for covariates

Additionally, one of the reviewers suggested not adjusting for BMI in the PheWAS analysis, given that BMI is known to be causally linked to urate levels. If adjusting for BMI, why not adjust for renal function?

It is acknowledged that using genetic variants as proxy of exposure has the advantage of not be influenced by most confounding factors, since genetic variants are fixed at conception and typically do not change due to environment factors. Adjusting with covariates may lead to bias in the causal estimate when a covariate is on the causal pathway of the exposure to the outcome or is a collider or causally downstream of a collider (565). Briefly, a collider is variable that could be influenced both by the exposure and by the outcome. This was the reason why we did not adjust for renal function, since it is possible that renal function could

be influenced by both the SUA level and multiple disease conditions (e.g., chronic kidney disease, or hypertensive nephropathy).

However, in some cases, adjustment for covariates is necessary to ensure validity of the IVs. One example is population stratification, in which the sample population consists of ethnic subgroups that have different distributions of the IVs and the outcome. An association between the IVs and outcome may be solely due to differences in ethnicity instead of any biological effect of the exposure. This can be addressed at least partially by adjusting for genetic PCs in the analysis. When I was performing the analysis, adjusting 5-6 principal components of ancestry was standard practice (in the past), thus I followed the old fashion and only adjusted for the first 5 principal components. However, recent results suggest that 40 PCs are more appropriate for large cohorts. This may present as a limitation of this study.

Besides, if the measured covariates can explain variation in the exposure or the outcome, then including such covariates in the analysis can generally improve the statistical power and increase the precision of the causal estimate. The rationale here is that accounting for a true risk factor decreases the residual variance of the outcome/phenotype and therefore increases the ratio of the true effect size of a predictor of interest (i.e. SUA) over the total phenotypic variance, which leads to increased statistical power and better precision in the causal estimate. Under these situations, adjusting for covariates is supportive.

In relation to BMI, it is unlikely that it is a collider, as BMI is known to be upstream of the SUA metabolism pathway, as well as a known causal factor for disease development. Besides, given that BMI is a common risk factor for a large number of diseases, it is likely that adjusting for BMI contributes to stronger statistical power of PheWAS by reducing the residual variance. With the purpose of being more rigorous, I re-performed the PheWAS without adjusting for BMI as a sensitivity analysis. The unadjusted results have been provided as supplementary material, in which all identified genetic associations retained their statistical significance with slight changes in their regression coefficients.

7.2.5.4 Case ascertainment

For the causal effect of SUA level on gout, one reviewer questioned that “regarding use of ICD codes for gout, they are likely to be un-representative of gout given that hospitalised gout is complicated by co-morbidities (566) - this should be explicitly investigated and acknowledged as a limitation. It would be interesting for the authors to use the classification criteria of Cadzow *et al* (480) to calculate a causal effect size for serum urate in gout.”

Considering the possibility of introducing information bias in this study, we did not incorporate the self-reported data into the PheWAS analysis and I agree with the reviewer that this is a limitation of this study, as acknowledged in *Chapter 5, Section "5.5 Discussion"*. Alternatively, we used a group of ICD diagnosis codes (ICD-10 code "M10" and its sub-codes; ICD-9 code "274" and its sub-codes) to represent gout. Gout cases were accordingly ascertained based on the primary or secondary hospital discharge coding. Gout cases derived from hospitalisation, as noted by the reviewer, are likely to be un-representative of gout given that hospitalised gout is complicated by co-morbidities (566). We are also interested whether using the classification criteria proposed by Cadzow *et al* (480) would make any change to the causal effect size of SUA on gout ascertained from different resources. By following the reviewer's suggestion, I conducted a sensitivity analysis to compare the MR estimates for hospital-diagnosed gout, self-reported gout and hospital-diagnosed/self-reported gout. The MR IVW/Egger estimates are consistently statistically significant in any of the cases, but the effect size are different and the self-reported gout has the largest relative risk.

7.3 Interpretation of the main findings

Respective findings and conclusions derived from the umbrella review, the MR-PheWAS of the interim release of UK Biobank data and the PWMR of the full UK Biobank data have been discussed in the corresponding chapters (i.e., ***Chapter 3, Chapter 5***, and ***Chapter 6***). In this section, I will provide an overall discussion and conclusion by incorporating findings from these three chapters.

7.3.1 Causality supported by convincing evidence

7.3.1.1 Gout

The causal relationship between uric acid and gout is robustly verified in this thesis with consistent evidence from the umbrella review, the MR-PheWAS of the interim release of UK Biobank data and the PWMR of the full UK Biobank data. Regarding the causal effect size on gout risk, estimates from these three sources of evidence are generally consistent. The umbrella review identified a MR analysis investigating the causal effect with 3,151 gout cases and 68,350 controls and reported that 1 SD increase in genetically determined SUA levels was associated with an increased risk of gout with an OR of 5.84 (95%CI: 4.56 to 7.49) (327). The causal effect size of SUA level on gout estimated form the MR-PheWAS including 1,003 gout cases and 119,555 controls presented a similar estimate with wider confidence interval (OR=4.58; 95%CI: 2.72 to 7.72) after correction of pleiotropy in MR

Egger regression (461). The causal effect estimated from the PWMR including 2,532 gout cases and 335,108 controls in UK Biobank reported a very similar magnitude of effect size on gout (OR=5.37; 95%CI: 4.67 to 6.18; p= 4.27×10^{-123}). Overall, the causal effect of SUA level on gout is consistently supported in both direction and magnitude of effect size by the three lines of evidence incorporated in this thesis.

7.3.2 Association supported by suggestive evidence

7.3.2.1 Hypertension

Apart from gout, the association between SUA and hypertension was classified as highly suggestive. In the umbrella review, a meta-analysis of observational studies provided highly suggestive evidence to support this association. The selected (largest) meta-analysis of RCTs on corresponding intermediate traits or surrogate outcomes (e.g., SBP, DBP) showed concordant evidence to support the causal effect. Evidence from published MR studies reported discordant evidence, in which the causal relationship was not verified (232). The MR-PheWAS and PWMR analysis performed in this thesis using data from UK Biobank demonstrated that SUA level shared genetic risk with hypertension at multiple loci, however, due to the presence of unbalanced pleiotropy detected by the MR Egger analysis, the causal association was not robustly inferred (461). Overall, when considering our emerging findings together with the previous evidence from umbrella review, it is reasonable to conclude an independent association between SUA level and hypertension, although there is not enough evidence at present to robustly conclude that this is causal.

7.3.2.2 Heart diseases

In the umbrella review, a wide range of heart diseases (including coronary heart disease, heart failure, and atrial fibrillation) has been identified to be associated with SUA level from observational studies, but there was a lack of concordance with clinically relevant endpoints from RCTs or surrogate endpoints from MR studies. Therefore, the evidence from the umbrella review is insufficient to support any casual effect of SUA level on these outcomes (232). In the MR-PheWAS and PWMR analysis, a wide range of cardiac diseases of varying severity, including coronary atherosclerosis, angina pectoris, ischaemic heart diseases, acute/old myocardial infarction and heart failure, were identified to be associated with the genetically determined SUA level (461). The association between SUA and cardiac diseases is favoured in both the previous studies and multiple analyses conducted in this thesis, however, given the same caveat in the causal inference with hypertension, a conclusion of causality on cardiac diseases is not robust enough.

7.3.2.3 Metabolic disorders

Epidemiological evidence from the umbrella review indicated that high SUA level was associated with some of the components of metabolic syndrome (537), such as blood glucose levels or type 2 diabetes. However, the MR-PheWAS and PWMR analysis in this thesis did not identify any significant association between SUA and specific metabolic other than hypercholesterolaemia. When taken together the line of evidence from experimental studies, which indicated that high SUA level promoted the oxidation of low-density lipoprotein and disrupted the process of reverse cholesterol transport, it is likely that there is an underlying biological link between hyperuricaemia and hypercholesterolaemia.

7.3.3 Association supported by weak evidence

7.3.3.1 Cerebrovascular diseases

The role of SUA level in the development of cerebrovascular diseases is debatable. The findings from this thesis are not completely consistent. Previous studies assessed in the umbrella review reported conflicting results for the association between SUA level and cerebrovascular diseases (e.g. stroke), where some studies suggested that high SUA level was neuroprotective whereas some studies reported that high SUA level was injurious. The TreeWAS analysis identified a positive association between the weighted GRS of SUA and cerebrovascular diseases, but this association was not replicated in the PheWAS analysis. In conclusion, findings in this thesis do not support the hypothesis that high SUA levels are neuroprotective as an anti-oxidant in the setting of cerebrovascular diseases.

7.3.3.2 Respiratory diseases

In addition to the findings discussed above, a few other disease outcomes were identified to be associated with SUA level with weak evidence. The sex-stratified MR-PheWAS analysis identified that a group of respiratory diseases were potentially linked with the genetically determined SUA level in women. The umbrella review did not provide any evidence regarding the respiratory diseases, but findings from experimental studies demonstrated that human airway epithelial cells and lung tissue expressed a functional UA production/secretion system and UA was crucial in mediating the development of allergic airway diseases and regulating the antigen-specific T-cell proliferation (494-497). Overall, the evidence in relation to the association between SUA level and respiratory diseases has not been well explored (499, 500). Our study contributes knowledge to the clinical relevance of SUA level in lung health and respiratory diseases.

7.3.3.3 Cataract

Another possible association identified from the PheWAS analysis was that of SUA level with cataract. Although the MR analyses did not indicate any putative causal links (with limited power = 0.37), a review of the literature demonstrated that various presentations of ocular abnormalities have been described in relation to SUA level. These include the depositions of MSU crystals in different locations in the eye (e.g. cornea, conjunctivae, and iris), and the abnormalities of dry eye syndrome and intraocular hypertension frequently observed in hyperuricaemia patients (567). In addition, an association between gout and nuclear, posterior subscapular and cortical cataracts have also been reported, in which gout was suggested as a risk factor for cataract development (568). The exact biological role of SUA level in ocular disease development has not been fully understood. Our observation adds to the current evidence indicating a potential relationship between SUA level and ocular diseases. We believe that this accumulating evidence should lead to further investigation of ocular diseases in patients with hyperuricaemia and gout if this is of clinical relevance.

8 CONCLUSION

This thesis presents a comprehensive investigation on the health outcomes in relation to SUA levels. In conclusion, I firstly performed an umbrella review to provide a comprehensive overview of reported associations between serum uric acid (SUA) levels and a wide range of health outcomes by incorporating evidence from systematic reviews and meta-analyses of observational studies, meta-analyses of randomised controlled trials, and Mendelian randomisation studies. This remarkable assembly of evidence explored 136 unique health outcomes and concluded convincing evidence of a causal clear role of SUA level in gout. I then investigated the associations of the 31 individual SNPs related to SUA level with a very wide range of disease outcomes by using MR-PheWAS design (phenome-wide association study incorporated with Mendelian randomisation [MR]) with the interim release data of UK Biobank (n=120,091); this MR-PheWAS analysis demonstrated that SUA level shared genetic risk loci with multiple disease outcomes, particularly cardiovascular/metabolic diseases and autoimmune disorders. When balancing out the pleiotropy on MR Egger analysis, a robust conclusion on causality was made only for gout. When enlarging the sample size of PheWAS with 3-fold more cases by using the data from the full UK Biobank data (n=339,256), the analysis demonstrated that genetically determined SUA level is independently and consistently associated with several disease groups including inflammatory polyarthropathies (e.g., gout), hypertensive disease (e.g., essential hypertension), ischaemic heart diseases (e.g., coronary atherosclerosis, myocardial infarction, chronic ischaemic heart disease), metabolic disorders (e.g., hypercholesterolaemia) and suggest possible association with cerebrovascular diseases (e.g., cerebral infarction). These associations with gout, CHD, myocardial infarction and decreased level of HDL-c were successfully replicated in different populations by analysing data from various GWAS consortia. The analysis of causal inference detected the existence of unbalanced genetic pleiotropy in most of the associations. To further investigate the influence of pleiotropy, I re-calculated the PheWAS estimates by using a number of GRSs created based on their association with a set of metabolic traits. The GRS of urate-specific loci was only associated with gout and its encompassing disease group of inflammatory polyarthropathies, but not with any cardiovascular/metabolic diseases. In contrast, the GRSs of pleiotropic loci on BMI, BP, lipids and glucose showed association with both gout and the cardiovascular/metabolic diseases. When removing any group of pleiotropic loci from the creation of GRS, their association with hypertensive diseases, heart diseases, and metabolic disorders were not

statistically significant. When balancing out the potential pleiotropic effects in Egger MR, a causal effect can only be verified for gout.

The causal inference in this study is also limited by the common difficulty of pleiotropy caused by the use of multiple genetic instruments. Although we have performed a number of sensitivity analyses by excluding the key pleiotropic loci, there is still the probability of unmeasured pleiotropy. In particular, unbalanced pleiotropy was recognised as an issue for the causal inference on the association between SUA level and hypertension. The potential causal link of SUA level with respiratory diseases and ocular diseases is also worthy of further investigation. Overall, when taken together the findings from umbrella review, MR-PheWAS, TreeWAS, MR replication and sensitivity analysis, I conclude that there are robust associations between urate and several disease groups, including gout, hypertensive diseases, heart diseases and metabolic disorders of lipids, but the causal role of urate only exists in gout. The study in this thesis indicates that the pleiotropic effects of genetic variants on urate and metabolic traits may contribute to the observed associations between urate and cardiovascular/metabolic diseases. These findings suggest that urate could be a good predictor for the cardiovascular/metabolic disease risk, but may not a therapeutic target for reducing the risk of cardiovascular/metabolic diseases. Further investigation on therapies targeting on a more distal mediator (or biological pathway) shared between urate and metabolic traits would be beneficial for the treatment of gout and the primary prevention of cardiovascular/metabolic comorbidities.

The research work presented in this thesis provides a comprehensive and thorough examination of urate-associated disease outcomes across the whole ICD spectrum and illustrates how serum urate level might influence the overall health. The detection of a multitude of cross-phenotype associations in this study adds to our understanding of the extent of shared genetic/biological components between urate and metabolic traits and increase our knowledge of how the role of urate should be interpreted and used in clinical practice in the management of related disease conditions. The overall analyses were limited by the lack of SUA biomarker data in UK Biobank. As UK Biobank is currently conducting biomarker assays, it would be beneficial to assess whether measured SUA level, rather than its genetic proxies, is also associated with the observed disease outcomes.

With regard to the PheWAS methodology applied in this thesis, updating the PheCODE schema or applying a novel Bayesian analysis with a tree-structured phenotypic model progresses the establishment of a framework or workflow of PheWAS design for comprehensively interrogating the clinical influence of a biomarker. The ongoing challenge

for PheWAS is the possibility of misclassification, misdiagnosis and miscoding in routine healthcare data. It is expected that exploring the genetic basis of the healthcare phenome can expose disease areas where improvements are required to ameliorate disease perception or strengthen diagnostic practices. Digital phenotyping using genetic data together with longitudinal clinical records, physical measures, images and biomarkers may be helpful to rectify misclassification, misdiagnosis and miscoding present in healthcare data and to infer missing phenotypes. Processing linkage to general practice records and out-patient data would also be helpful to build up the individual's phenome based on widely-covered and accurately-defined criteria in the future.

REFERENCES

1. McCrudden FH. Uric Acid. BiblioBazaar. 2008.
2. Watanabe S, Kang DH, Feng L, Nakagawa T, Kanellis J, Lan H, et al. Uric acid, hominoid evolution, and the pathogenesis of salt-sensitivity. Hypertension (Dallas, Tex : 1979). 2002;40(3):355-60.
3. Alvarez-Lario B, Macarron-Vicente J. Uric acid and evolution. Rheumatology (Oxford, England). 2010;49(11):2010-5.
4. Campion DS, Olsen R, Bluestone R, Klinenberg JR. Binding of urate by serum proteins. Arthritis and rheumatism. 1975;18(6 Suppl):747-9.
5. Martillo MA, Nazzal L, Crittenden DB. The crystallization of monosodium urate. Current rheumatology reports. 2014;16(2):400.
6. Jin M, Yang F, Yang I, Yin Y, Luo JJ, Wang H, et al. Uric acid, hyperuricemia and vascular diseases. Frontiers in bioscience (Landmark edition). 2012;17:656-69.
7. Scott JT, Holloway VP, Glass HI, Arnot RN. Studies of uric acid pool size and turnover rate. Annals of the rheumatic diseases. 1969;28(4):366-73.
8. Maiuolo J, Oppedisano F, Gratteri S, Muscoli C, Mollace V. Regulation of uric acid metabolism and excretion. International journal of cardiology. 2016;213:8-14.
9. Pang B, McFaline JL, Burgis NE, Dong M, Taghizadeh K, Sullivan MR, et al. Defects in purine nucleotide metabolism lead to substantial incorporation of xanthine and hypoxanthine into DNA and RNA. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(7):2319-24.
10. Rosemeyer H. The chemodiversity of purine as a constituent of natural products. Chemistry & biodiversity. 2004;1(3):361-401.
11. Moffatt BA, Ashihara H. Purine and pyrimidine nucleotide synthesis and metabolism. The arabidopsis book. 2002;1:e0018.
12. Peifer S, Barduhn T, Zimmet S, Volmer DA, Heinzle E, Schneider K. Metabolic engineering of the purine biosynthetic pathway in Corynebacterium glutamicum results in increased intracellular pool sizes of IMP and hypoxanthine. Microbial cell factories. 2012;11:138.
13. Ichida K, Amaya Y, Okamoto K, Nishino T. Mutations associated with functional disorder of xanthine oxidoreductase and hereditary xanthinuria in humans. International journal of molecular sciences. 2012;13(11):15475-95.
14. Jurecka A. Inborn errors of purine and pyrimidine metabolism. Journal of inherited metabolic disease. 2009;32(2):247-63.
15. Sorensen LB, Levinson DJ. Origin and extrarenal elimination of uric acid in man. Nephron. 1975;14(1):7-20.
16. Xu X, Li C, Zhou P, Jiang T. Uric acid transporters hiding in the intestine. Pharmaceutical biology. 2016;54(12):3151-5.
17. Hosomi A, Nakanishi T, Fujita T, Tamai I. Extra-renal elimination of uric acid via intestinal efflux transporter BCRP/ABCG2. PloS one. 2012;7(2):e30456.
18. Rouf MA, Lomprey RF, Jr. Degradation of uric acid by certain aerobic bacteria. Journal of bacteriology. 1968;96(3):617-22.
19. Braun EJ, Campbell CE. Uric acid decomposition in the lower gastrointestinal tract. The Journal of experimental zoology Supplement : published under auspices of the

American Society of Zoologists and the Division of Comparative Physiology and Biochemistry. 1989;3:70-4.
20. Wyngaarden JB, Stetten D, Jr. Uricolysis in normal man. The Journal of biological chemistry. 1953;203(1):9-21.
21. Lipkowitz MS. Regulation of uric acid excretion by the kidney. Current rheumatology reports. 2012;14(2):179-88.
22. Bobulescu IA, Moe OW. Renal transport of uric acid: evolving concepts and uncertainties. Advances in chronic kidney disease. 2012;19(6):358-71.
23. Enomoto A, Kimura H, Chairoungdua A, Shigeta Y, Jutabha P, Cha SH, et al. Molecular identification of a renal urate anion exchanger that regulates blood urate levels. Nature. 2002;417(6887):447-52.
24. Augustin R, Carayannopoulos MO, Dowd LO, Phay JE, Moley JF, Moley KH. Identification and characterization of human glucose transporter-like protein-9 (GLUT9): alternative splicing alters trafficking. The Journal of biological chemistry. 2004;279(16):16229-36.
25. Phay JE, Hussain HB, Moley JF. Cloning and expression analysis of a novel member of the facilitative glucose transporter family, SLC2A9 (GLUT9). Genomics. 2000;66(2):217-20.
26. Huls M, Brown CD, Windass AS, Sayer R, van den Heuvel JJ, Heemskerk S, et al. The breast cancer resistance protein transporter ABCG2 is expressed in the human kidney proximal tubule apical membrane. Kidney international. 2008;73(2):220-5.
27. Lee K, Belinsky MG, Bell DW, Testa JR, Kruh GD. Isolation of MOAT-B, a widely expressed multidrug resistance-associated protein/canalicular multispecific organic anion transporter-related transporter. Cancer research. 1998;58(13):2741-7.
28. van Aubel RA, Smeets PH, Peters JG, Bindels RJ, Russel FG. The MRP4/ABCC4 gene encodes a novel apical organic anion transporter in human kidney proximal tubules: putative efflux pump for urinary cAMP and cGMP. Journal of the American Society of Nephrology : JASN. 2002;13(3):595-603.
29. Culleton BF. Uric acid and cardiovascular disease: a renal-cardiac relationship? Current opinion in nephrology and hypertension. 2001;10(3):371-5.
30. Reyes AJ, Leary WP. The increase in serum uric acid induced by diuretics could be beneficial to cardiovascular prognosis in hypertension: a hypothesis. Journal of hypertension. 2003;21(9):1775-7.
31. Glantzounis GK, Tsimoyiannis EC, Kappas AM, Galaris DA. Uric acid and oxidative stress. Current pharmaceutical design. 2005;11(32):4145-51.
32. Sautin YY, Johnson RJ. Uric acid: the oxidant-antioxidant paradox. Nucleosides, nucleotides & nucleic acids. 2008;27(6):608-19.
33. Sautin YY, Nakagawa T, Zharikov S, Johnson RJ. Adverse effects of the classic antioxidant uric acid in adipocytes: NADPH oxidase-mediated oxidative/nitrosative stress. American journal of physiology Cell physiology. 2007;293(2):C584-96.
34. Oda M, Satta Y, Takenaka O, Takahata N. Loss of urate oxidase activity in hominoids and its evolutionary implications. Molecular biology and evolution. 2002;19(5):640-53.
35. Ames BN, Cathcart R, Schwiers E, Hochstein P. Uric acid provides an antioxidant defense in humans against oxidant- and radical-caused aging and cancer: a hypothesis.

Proceedings of the National Academy of Sciences of the United States of America. 1981;78(11):6858-62.
36. Kellogg EW, 3rd, Fridovich I. Liposome oxidation and erythrocyte lysis by enzymically generated superoxide and hydrogen peroxide. The Journal of biological chemistry. 1977;252(19):6721-8.
37. Becker BF. Towards the physiological function of uric acid. Free radical biology & medicine. 1993;14(6):615-31.
38. Kuzkaya N, Weissmann N, Harrison DG, Dikalov S. Interactions of peroxynitrite with uric acid in the presence of ascorbate and thiols: implications for uncoupling endothelial nitric oxide synthase. Biochemical pharmacology. 2005;70(3):343-54.
39. Robinson KM, Morre JT, Beckman JS. Triuret: a novel product of peroxynitrite-mediated oxidation of urate. Archives of biochemistry and biophysics. 2004;423(1):213-7.
40. Suzuki T. Nitrosation of uric acid induced by nitric oxide under aerobic conditions. Nitric oxide : biology and chemistry. 2007;16(2):266-73.
41. Waugh WH. Inhibition of iron-catalyzed oxidations by attainable uric acid and ascorbic acid levels: therapeutic implications for Alzheimer's disease and late cognitive impairment. Gerontology. 2008;54(4):238-43.
42. Whiteman M, Ketsawatsakul U, Halliwell B. A reassessment of the peroxynitrite scavenging activity of uric acid. Annals of the New York Academy of Sciences. 2002;962:242-59.
43. Muraoka S, Miura T. Inhibition by uric acid of free radicals that damage biological molecules. Pharmacology & toxicology. 2003;93(6):284-9.
44. Pham-Huy LA, He H, Pham-Huy C. Free radicals, antioxidants in disease and health. International journal of biomedical science : IJBS. 2008;4(2):89-96.
45. Maples KR, Mason RP. Free radical metabolite of uric acid. The Journal of biological chemistry. 1988;263(4):1709-12.
46. Santos CX, Anjos EI, Augusto O. Uric acid oxidation by peroxynitrite: multiple reactions, free radical formation, and amplification of lipid oxidation. Archives of biochemistry and biophysics. 1999;372(2):285-94.
47. Chung HY, Baek BS, Song SH, Kim MS, Huh JI, Shim KH, et al. Xanthine dehydrogenase/xanthine oxidase and oxidative stress. Age. 1997;20(3):127-40.
48. Waud WR, Rajagopalan KV. Purification and properties of the NAD+-dependent (type D) and O2-dependent (type O) forms of rat liver xanthine dehydrogenase. Archives of biochemistry and biophysics. 1976;172(2):354-64.
49. Higgins P, Dawson J, Lees KR, McArthur K, Quinn TJ, Walters MR. Xanthine oxidase inhibition for the treatment of cardiovascular disease: a systematic review and meta-analysis. Cardiovascular therapeutics. 2012;30(4):217-26.
50. Kanbay M, Siriopol D, Nistor I, Elcioglu OC, Telci O, Takir M, et al. Effects of allopurinol on endothelial dysfunction: a meta-analysis. American journal of nephrology. 2014;39(4):348-56.
51. Hall AP, Barry PE, Dawber TR, McNamara PM. Epidemiology of gout and hyperuricemia. A long-term population study. The American journal of medicine. 1967;42(1):27-37.
52. Mandal AK, Mount DB. The molecular physiology of uric acid homeostasis. Annual review of physiology. 2015;77:323-45.

53. Bardin T, Richette P. Definition of hyperuricemia and gouty conditions. Current opinion in rheumatology. 2014;26(2):186-91.
54. Desideri G, Castaldo G, Lombardi A, Mussap M, Testa A, Pontremoli R, et al. Is it time to revise the normal range of serum uric acid levels? European review for medical and pharmacological sciences. 2014;18(9):1295-306.
55. Richette P, Doherty M, Pascual E, Barskova V, Becce F, Castaneda-Sanabria J, et al. 2016 updated EULAR evidence-based recommendations for the management of gout. Annals of the rheumatic diseases. 2017;76(1):29-42.
56. Stamp L, Dalbeth N. Urate-lowering therapy for asymptomatic hyperuricaemia: A need for caution. Seminars in arthritis and rheumatism. 2016.
57. Balasubramaniam S, Duley JA, Christodoulou J. Inborn errors of purine metabolism: clinical update and therapies. Journal of inherited metabolic disease. 2014;37(5):669-86.
58. Nyhan WL, O'Neill JP, Jinnah HA, Harris JC. Lesch-Nyhan Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mefford HC, et al., editors. GeneReviews((R)). Seattle (WA): University of Washington, Seattle

University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.; 1993.
59. Kelley WN, Rosenbloom FM, Henderson JF, Seegmiller JE. A specific enzyme defect in gout associated with overproduction of uric acid. Proceedings of the National Academy of Sciences of the United States of America. 1967;57(6):1735-9.
60. de Brouwer AP, van Bokhoven H, Nabuurs SB, Arts WF, Christodoulou J, Duley J. PRPS1 mutations: four distinct syndromes and potential treatment. American journal of human genetics. 2010;86(4):506-18.
61. Chou JY, Matern D, Mansfield BC, Chen YT. Type I glycogen storage diseases: disorders of the glucose-6-phosphatase complex. Current molecular medicine. 2002;2(2):121-43.
62. Davidson MB, Thakkar S, Hix JK, Bhandarkar ND, Wong A, Schreiber MJ. Pathophysiology, clinical consequences, and treatment of tumor lysis syndrome. The American journal of medicine. 2004;116(8):546-54.
63. Hare JM, Johnson RJ. Uric acid predicts clinical outcomes in heart failure: insights regarding the role of xanthine oxidase and uric acid in disease pathophysiology. Circulation. 2003;107(15):1951-3.
64. Jefferson JA, Escudero E, Hurtado ME, Kelly JP, Swenson ER, Wener MH, et al. Hyperuricemia, hypertension, and proteinuria associated with high-altitude polycythemia. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2002;39(6):1135-42.
65. Torralba KD, De Jesus E, Rachabattula S. The interplay between diet, urate transporters and the risk for gout and hyperuricemia: current and future directions. International journal of rheumatic diseases. 2012;15(6):499-506.
66. Faller J, Fox IH. Ethanol-induced hyperuricemia: evidence for increased urate production by activation of adenine nucleotide turnover. The New England journal of medicine. 1982;307(26):1598-602.
67. Lieber CS, Jones DP, Losowsky MS, Davidson CS. Interrelation of uric acid and ethanol metabolism in man. The Journal of clinical investigation. 1962;41:1863-70.

68. Emmerson BT. Effect of oral fructose on urate production. Annals of the rheumatic diseases. 1974;33(3):276-80.
69. Johnson RJ, Segal MS, Sautin Y, Nakagawa T, Feig DI, Kang DH, et al. Potential role of sugar (fructose) in the epidemic of hypertension, obesity and the metabolic syndrome, diabetes, kidney disease, and cardiovascular disease. The American journal of clinical nutrition. 2007;86(4):899-906.
70. Lloyd-Mostyn RH, Lord PS, Glover R, West C, Gilliland IC. Uric acid metabolism in starvation. Annals of the rheumatic diseases. 1970;29(5):553-5.
71. Gao B, Zhou J, Ge J, Zhang Y, Chen F, Lau WB, et al. Association of maximum weight with hyperuricemia risk: a retrospective study of 21,414 Chinese people. PloS one. 2012;7(11):e51186.
72. Fox IH, Halperin ML, Goldstein MB, Marliss ER. Renal excretion of uric acid during prolonged fasting. Metabolism: clinical and experimental. 1976;25(5):551-9.
73. Green HJ, Fraser IG. Differential effects of exercise intensity on serum uric acid concentration. Medicine and science in sports and exercise. 1988;20(1):55-9.
74. Chou CT, Lai JS. The epidemiology of hyperuricaemia and gout in Taiwan aborigines. British journal of rheumatology. 1998;37(3):258-62.
75. Li Y, Stamler J, Xiao Z, Folsom A, Tao S, Zhang H. Serum uric acid and its correlates in Chinese adult populations, urban and rural, of Beijing. The PRC-USA Collaborative Study in Cardiovascular and Cardiopulmonary Epidemiology. International journal of epidemiology. 1997;26(2):288-96.
76. Nan H, Qiao Q, Dong Y, Gao W, Tang B, Qian R, et al. The prevalence of hyperuricemia in a population of the coastal city of Qingdao, China. The Journal of rheumatology. 2006;33(7):1346-50.
77. Li D, Yu X, Zhou X, Siriamornpun S, Wahlqvist ML. Uric acid status and its correlates in Hangzhou urban population. Asia Pacific journal of clinical nutrition. 2006;15(1):102-6.
78. You L, Liu A, Wuyun G, Wu H, Wang P. Prevalence of hyperuricemia and the relationship between serum uric acid and metabolic syndrome in the Asian Mongolian area. Journal of atherosclerosis and thrombosis. 2014;21(4):355-65.
79. Chang HY, Pan WH, Yeh WT, Tsai KS. Hyperuricemia and gout in Taiwan: results from the Nutritional and Health Survey in Taiwan (1993-96). The Journal of rheumatology. 2001;28(7):1640-6.
80. Nagahama K, Iseki K, Inoue T, Touma T, Ikemiya Y, Takishita S. Hyperuricemia and cardiovascular risk factor clustering in a screened cohort in Okinawa, Japan. Hypertension research : official journal of the Japanese Society of Hypertension. 2004;27(4):227-33.
81. Darmawan J, Valkenburg HA, Muirden KD, Wigley RD. The epidemiology of gout and hyperuricemia in a rural population of Java. The Journal of rheumatology. 1992;19(10):1595-9.
82. Ryu S, Chang Y, Zhang Y, Kim SG, Cho J, Son HJ, et al. A cohort study of hyperuricemia in middle-aged South Korean men. American journal of epidemiology. 2012;175(2):133-43.
83. Prior IA, Rose BS. Uric acid, gout and public health in the South Pacific. The New Zealand medical journal. 1966;65(405):295-300.

84. Jackson L, Taylor R, Faaiuso S, Ainuu SP, Whitehouse S, Zimmet P. Hyperuricaemia and gout in Western Samoans. Journal of chronic diseases. 1981;34(2-3):65-75.
85. Tuomilehto J, Zimmet P, Wolf E, Taylor R, Ram P, King H. Plasma uric acid level and its association with diabetes mellitus and some biologic parameters in a biracial population of Fiji. American journal of epidemiology. 1988;127(2):321-36.
86. Nabipour I, Sambrook PN, Blyth FM, Janu MR, Waite LM, Naganathan V, et al. Serum uric acid is associated with bone health in older men: a cross-sectional population-based study. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2011;26(5):955-64.
87. Brennan PJ, Simpson JM, McGilchrist CA, Black BR. Serum lipids and other coronary risk factors in Sydney blood donors: a comparison with other Australian subjects. Australian and New Zealand Journal of Public Health. 1980;4(1):1-6.
88. Stamp LK, Wells JE, Pitama S, Faatoese A, Doughty RN, Whalley G, et al. Hyperuricaemia and gout in New Zealand rural and urban Māori and non-Māori communities. Internal medicine journal. 2013;43(6):678-84.
89. Brauer GW, Prior IA. A prospective study of gout in New Zealand Maoris. Annals of the rheumatic diseases. 1978;37(5):466-72.
90. Lohsoonthorn V, Dhanamun B, Williams MA. Prevalence of hyperuricemia and its relationship with metabolic syndrome in Thai adults receiving annual health exams. Archives of medical research. 2006;37(7):883-9.
91. Zhu Y, Pandya BJ, Choi HK. Prevalence of gout and hyperuricemia in the US general population: the National Health and Nutrition Examination Survey 2007-2008. Arthritis and rheumatism. 2011;63(10):3136-41.
92. Mikkelsen WM, Dodge HJ, Valkenburg H. The distribution of serum uric acid values in a population unselected as to gout or hyperuricemia: Tecumseh, Michigan 1959-1960. The American journal of medicine. 1965;39:242-51.
93. O'Brien WM, Burch TA, Bunim JJ. Genetics of hyperuricaemia in Blackfeet and Pima Indians. Annals of the rheumatic diseases. 1966;25(2):117-9.
94. Lopez-Molina R, Parra-Cabrera S, Lopez-Ridaura R, Gonzalez-Villalpando ME, Ferrannini E, Gonzalez-Villalpando C. Sweetened beverages intake, hyperuricemia and metabolic syndrome: the Mexico City Diabetes Study. Salud publica de Mexico. 2013;55(6):557-63.
95. Tavares EF, Vieira-Filho JP, Andriolo A, Sanudo A, Gimeno SG, Franco LJ. Metabolic profile and cardiovascular risk patterns of an Indian tribe living in the Amazon Region of Brazil. Human biology. 2003;75(1):31-46.
96. Gimeno SG, Rodrigues D, Cano EN, Lima EE, Schaper M, Pagliaro H, et al. Cardiovascular risk factors among Brazilian Karib indigenous peoples: Upper Xingu, Central Brazil, 2000-3. Journal of epidemiology and community health. 2009;63(4):299-304.
97. Al-Arfaj AS. Hyperuricemia in Saudi Arabia. Rheumatology international. 2001;20(2):61-4.
98. Sturge RA, Scott JT, Kennedy AC, Hart DP, Buchanan WW. Serum uric acid in England and Scotland. Annals of the rheumatic diseases. 1977;36(5):420-7.
99. Badley EM, Meyrick JS, Wood PH. Gout and serum uric acid levels in the Cotswolds. Rheumatology and rehabilitation. 1978;17(3):133-42.

100. Reunanen A, Takkunen H, Knekt P, Aromaa A. Hyperuricemia as a risk factor for cardiovascular mortality. Acta medica Scandinavica Supplementum. 1982;668:49-59.
101. Trifiro G, Morabito P, Cavagna L, Ferrajolo C, Pecchioli S, Simonetti M, et al. Epidemiology of gout and hyperuricaemia in Italy during the years 2005-2009: a nationwide population-based study. Annals of the rheumatic diseases. 2013;72(5):694-700.
102. Zalokar J, Lellouch J, Claude JR, Kuntz D. Serum uric acid in 23,923 men and gout in a subsample of 4257 men in France. Journal of chronic diseases. 1972;25(5):305-12.
103. Sari I, Akar S, Pakoz B, Sisman AR, Gurler O, Birlik M, et al. Hyperuricemia and its related factors in an urban population, Izmir, Turkey. Rheumatology international. 2009;29(8):869-74.
104. S. A. Shalnova, A. D. Deev, G. V. Artamonov, D. V. Duplyakov, A. Yu. Efanov, Yu. V. Zhernakova, et al. Hyperuricemia and its correlates in the Russian population. Rational pharmacotherapy in cardiology. 2014;10(2):153-9.
105. M.A. Makusidi, A Chijioke, P.M. Kolo, H.M. Liman, M.B. AbdulRahman, B Sani, et al. Prevalence and pattern of hyperuricemia in a survey among inhabitants of Sokoto metropolis, north western Nigeria. Research Journal of Health Sciences. 2016;4(1).
106. Moulin SR, Baldo MP, Souza JB, Luchi WM, Capingana DP, Magalhaes P, et al. Distribution of Serum Uric Acid in Black Africans and Its Association With Cardiovascular Risk Factors. Journal of clinical hypertension (Greenwich, Conn). 2017;19(1):45-50.
107. Conen D, Wietlisbach V, Bovet P, Shamlaye C, Riesen W, Paccaud F, et al. Prevalence of hyperuricemia and relation of serum uric acid with cardiovascular risk factors in a developing country. BMC public health. 2004;4:9.
108. Wallace KL, Riedel AA, Joseph-Ridge N, Wortmann R. Increasing prevalence of gout and hyperuricemia over 10 years among older adults in a managed care population. The Journal of rheumatology. 2004;31(8):1582-7.
109. Hakoda M. Recent trends in hyperuricemia and gout in Japan. Japan Medical Association journal : JMAJ. 2012;55(4):319-23.
110. Kuo CF, Grainge MJ, Zhang W, Doherty M. Global epidemiology of gout: prevalence, incidence and risk factors. Nature reviews Rheumatology. 2015;11(11):649-62.
111. Arromdee E, Michet CJ, Crowson CS, O'Fallon WM, Gabriel SE. Epidemiology of gout: is the incidence rising? The Journal of rheumatology. 2002;29(11):2403-6.
112. Juraschek SP, Miller ER, 3rd, Gelber AC. Body mass index, obesity, and prevalent gout in the United States in 1988-1994 and 2007-2010. Arthritis care & research. 2013;65(1):127-32.
113. Mikuls TR, Farrar JT, Bilker WB, Fernandes S, Schumacher HR, Jr., Saag KG. Gout epidemiology: results from the UK General Practice Research Database, 1990-1999. Annals of the rheumatic diseases. 2005;64(2):267-72.
114. Elliot AJ, Cross KW, Fleming DM. Seasonality and trends in the incidence and prevalence of gout in England and Wales 1994-2007. Annals of the rheumatic diseases. 2009;68(11):1728-33.
115. Klemp P, Stansfield SA, Castle B, Robertson MC. Gout is on the increase in New Zealand. Annals of the rheumatic diseases. 1997;56(1):22-6.
116. Miao Z, Li C, Chen Y, Zhao S, Wang Y, Wang Z, et al. Dietary and lifestyle changes associated with high prevalence of hyperuricemia and gout in the Shandong coastal cities of Eastern China. The Journal of rheumatology. 2008;35(9):1859-64.

117. Zeng Q, Wang Q, Chen R, Xiao Z, Huang S, Xu J. Primary gout in Shantou: a clinical and epidemiological study. Chinese medical journal. 2003;116(1):66-9.
118. Ford ES, Li C, Cook S, Choi HK. Serum concentrations of uric acid and the metabolic syndrome among US children and adolescents. Circulation. 2007;115(19):2526-32.
119. Hochberg MC, Smolen JS, ME W. Rheumatology 3 edition. New York: Mosby. 2003.
120. Smith EU, Diaz-Torne C, Perez-Ruiz F, March LM. Epidemiology of gout: an update. Best practice & research Clinical rheumatology. 2010;24(6):811-27.
121. Gaffo AL, Jacobs DR, Jr., Lewis CE, Mikuls TR, Saag KG. Association between being African-American, serum urate levels and the risk of developing hyperuricemia: findings from the Coronary Artery Risk Development in Young Adults cohort. Arthritis research & therapy. 2012;14(1):R4.
122. Portis AJ, Laliberte M, Tatman P, Moua M, Culhane-Pera K, Maalouf NM, et al. High prevalence of gouty arthritis among the Hmong population in Minnesota. Arthritis care & research. 2010;62(10):1386-91.
123. Prasad P, Krishnan E. Filipino gout: a review. Arthritis care & research. 2014;66(3):337-43.
124. Khanna D, Fitzgerald JD, Khanna PP, Bae S, Singh MK, Neogi T, et al. 2012 American College of Rheumatology guidelines for management of gout. Part 1: systematic nonpharmacologic and pharmacologic therapeutic approaches to hyperuricemia. Arthritis care & research. 2012;64(10):1431-46.
125. Roddy E, Choi HK. Epidemiology of gout. Rheumatic diseases clinics of North America. 2014;40(2):155-75.
126. Simkin PA. Urate excretion in normal and gouty men. Advances in experimental medicine and biology. 1977;76b:41-5.
127. Ramsdell CM, Kelley WN. The clinical significance of hypouricemia. Annals of internal medicine. 1973;78(2):239-42.
128. Eckel RH, Jakicic JM, Ard JD, de Jesus JM, Houston Miller N, Hubbard VS, et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25 Suppl 2):S76-99.
129. Choi HK, Liu S, Curhan G. Intake of purine-rich foods, protein, and dairy products and relationship to serum levels of uric acid: the Third National Health and Nutrition Examination Survey. Arthritis and rheumatism. 2005;52(1):283-9.
130. Choi HK, Soriano LC, Zhang Y, Rodriguez LA. Antihypertensive drugs and risk of incident gout among patients with hypertension: population based case-control study. BMJ. 2012;344:d8190.
131. Takahashi S, Moriwaki Y, Yamamoto T, Tsutsumi Z, Ka T, Fukuchi M. Effects of combination treatment using anti-hyperuricaemic agents with fenofibrate and/or losartan on uric acid metabolism. Annals of the rheumatic diseases. 2003;62(6):572-5.
132. Kenny JE, Goldfarb DS. Update on the pathophysiology and management of uric acid renal stones. Current rheumatology reports. 2010;12(2):125-9.
133. Coiffier B, Altman A, Pui CH, Younes A, Cairo MS. Guidelines for the management of pediatric and adult tumor lysis syndrome: an evidence-based review. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(16):2767-78.

134. Zhang W, Doherty M, Bardin T, Pascual E, Barskova V, Conaghan P, et al. EULAR evidence based recommendations for gout. Part II: Management. Report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Annals of the rheumatic diseases. 2006;65(10):1312-24.
135. Grassi D, Pontremoli R, Bocale R, Ferri C, Desideri G. Therapeutic approaches to chronic hyperuricemia and gout. High blood pressure & cardiovascular prevention : the official journal of the Italian Society of Hypertension. 2014;21(4):243-50.
136. Hande KR, Noone RM, Stone WJ. Severe allopurinol toxicity. Description and guidelines for prevention in patients with renal insufficiency. The American journal of medicine. 1984;76(1):47-56.
137. Rundles RW, Metz EN, Silberman HR. Allopurinol in the treatment of gout. Annals of internal medicine. 1966;64(2):229-58.
138. Singer JZ, Wallace SL. The allopurinol hypersensitivity syndrome. Unnecessary morbidity and mortality. Arthritis and rheumatism. 1986;29(1):82-7.
139. Terkeltaub R, Bushinsky DA, Becker MA. Recent developments in our understanding of the renal basis of hyperuricemia and the development of novel antihyperuricemic therapeutics. Arthritis research & therapy. 2006;8 Suppl 1:S4.
140. Becker MA, Schumacher HR, Jr., Wortmann RL, MacDonald PA, Palo WA, Eustace D, et al. Febuxostat, a novel nonpurine selective inhibitor of xanthine oxidase: a twenty-eight-day, multicenter, phase II, randomized, double-blind, placebo-controlled, dose-response clinical trial examining safety and efficacy in patients with gout. Arthritis and rheumatism. 2005;52(3):916-23.
141. Bohm M, Vuppalanchi R, Chalasani N. Febuxostat-induced acute liver injury. Hepatology (Baltimore, Md). 2016;63(3):1047-9.
142. Becker MA, Schumacher HR, Jr., Wortmann RL, MacDonald PA, Eustace D, Palo WA, et al. Febuxostat compared with allopurinol in patients with hyperuricemia and gout. The New England journal of medicine. 2005;353(23):2450-61.
143. Diaz-Torne C, Perez-Herrero N, Perez-Ruiz F. New medications in development for the treatment of hyperuricemia of gout. Current opinion in rheumatology. 2015;27(2):164-9.
144. Fam AG. Difficult gout and new approaches for control of hyperuricemia in the allopurinol-allergic patient. Current rheumatology reports. 2001;3(1):29-35.
145. Perez-Ruiz F, Sundy JS, Miner JN, Cravets M, Storgard C. Lesinurad in combination with allopurinol: results of a phase 2, randomised, double-blind study in patients with gout with an inadequate response to allopurinol. Annals of the rheumatic diseases. 2016;75(6):1074-80.
146. Burns CM, Wortmann RL. Latest evidence on gout management: what the clinician needs to know. Therapeutic advances in chronic disease. 2012;3(6):271-86.
147. Sundy JS, Baraf HS, Yood RA, Edwards NL, Gutierrez-Urena SR, Treadwell EL, et al. Efficacy and tolerability of pegloticase for the treatment of chronic gout in patients refractory to conventional treatment: two randomized controlled trials. Jama. 2011;306(7):711-20.
148. Alakel N, Middeke JM, Schetelig J, Bornhauser M. Prevention and treatment of tumor lysis syndrome, and the efficacy and role of rasburicase. OncoTargets and therapy. 2017;10:597-605.

149. Nath SD, Voruganti VS, Arar NH, Thameem F, Lopez-Alvarenga JC, Bauer R, et al. Genome scan for determinants of serum uric acid variability. Journal of the American Society of Nephrology : JASN. 2007;18(12):3156-63.
150. Kolz M, Johnson T, Sanna S, Teumer A, Vitart V, Perola M, et al. Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. PLoS genetics. 2009;5(6):e1000504.
151. Kottgen A, Albrecht E, Teumer A, Vitart V, Krumsiek J, Hundertmark C, et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nature genetics. 2013;45(2):145-54.
152. National Center of Biotechnology Information. Entrez Gene Summary for SLC22A12 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/116085.
153. Crown Human Genome Center. GeneCards Summary for SLC22A12 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC22A12.
154. Manolescu AR, Augustin R, Moley K, Cheeseman C. A highly conserved hydrophobic motif in the exofacial vestibule of fructose transporting SLC2A proteins acts as a critical determinant of their substrate selectivity. Molecular membrane biology. 2007;24(5-6):455-63.
155. National Center of Biotechnology Information. Entrez Gene Summary for SLC2A9 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/56606.
156. Center CHG. GeneCards Summary for SLC2A9 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC2A9.
157. National Center of Biotechnology Information. Entrez Gene Summary for ABCG2 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/9429.
158. Vlaming ML, Lagas JS, Schinkel AH. Physiological and pharmacological roles of ABCG2 (BCRP): recent findings in Abcg2 knockout mice. Advanced drug delivery reviews. 2009;61(1):14-25.
159. Center CHG. GeneCards Summary for ABCG2 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=ABCG2.
160. National Center of Biotechnology Information. Entrez Gene Summary for SLC22A11 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/55867.
161. Zhou F, Xu W, Hong M, Pan Z, Sinko PJ, Ma J, et al. The role of N-linked glycosylation in protein folding, membrane targeting, and substrate binding of human organic anion transporter hOAT4. Molecular pharmacology. 2005;67(3):868-76.
162. Center CHG. GeneCards Summary for SLC22A11 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC22A11.
163. National Center of Biotechnology Information. Entrez Gene Summary for SLC17A1 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/6568.
164. Chong SS, Kristjansson K, Zoghbi HY, Hughes MR. Molecular cloning of the cDNA encoding a human renal sodium phosphate transport protein and its assignment to chromosome 6p21.3-p23. Genomics. 1993;18(2):355-9.

165. Center CHG. GeneCards Summary for SLC17A1 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC22A11.
166. National Center of Biotechnology Information. Entrez Gene Summary for PDZK1 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/5174.
167. Kocher O, Comella N, Tognazzi K, Brown LF. Identification and partial characterization of PDZK1: a novel protein containing PDZ interaction domains. Laboratory investigation; a journal of technical methods and pathology. 1998;78(1):117-25.
168. Silver DL, Wang N, Vogel S. Identification of small PDZK1-associated protein, DD96/MAP17, as a regulator of PDZK1 and plasma high density lipoprotein levels. The Journal of biological chemistry. 2003;278(31):28528-32.
169. Center CHG. GeneCards Summary for PDZK1 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=PDZK1.
170. National Center of Biotechnology Information. Entrez Gene Summary for SLC16A9 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/220963.
171. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature. 2011;477(7362):54-60.
172. Center CHG. GeneCards Summary for SLC16A9 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC16A9.
173. Hayward BE, Dunlop N, Intody S, Leek JP, Markham AF, Warner JP, et al. Organization of the human glucokinase regulator gene GCKR. Genomics. 1998;49(1):137-42.
174. National Center of Biotechnology Information. Entrez Gene Summary for GCKR Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/2646.
175. Center CHG. GeneCards Summary for GCKR Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=GCKR.
176. National Center of Biotechnology Information. Entrez Gene Summary for INHBC Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/3626.
177. G Ht, Neidhardt H, Schneider C, Pohl J. Cloning of a new member of the TGF-beta family: a putative new activin beta C chain. Biochemical and biophysical research communications. 1995;206(2):608-13.
178. Center CHG. GeneCards Summary for INHBC Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=INHBC.
179. National Center of Biotechnology Information. Entrez Gene Summary for RREB1 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/6239.
180. Melani M, Simpson KJ, Brugge JS, Montell D. Regulation of cell adhesion and collective cell migration by hindsight and its human homolog RREB1. Current biology : CB. 2008;18(7):532-7.
181. Crown Human Genome Center. GeneCards Summary for RREB1 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=RREB1.

182. Drewes T, Senkel S, Holewa B, Ryffel GU. Human hepatocyte nuclear factor 4 isoforms are encoded by distinct and differentially expressed genes. Molecular and cellular biology. 1996;16(3):925-31.
183. National Center of Biotechnology Information. Entrez Gene Summary for HNF4G Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/3174.
184. Center CHG. GeneCards Summary for HNF4G Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=HNF4G.
185. National Center of Biotechnology Information. Entrez Gene Summary for SFMBT1 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/51460.
186. Morel S, Levy F, Burlet-Schiltz O, Brasseur F, Probst-Kepper M, Peitrequin AL, et al. Processing of some antigens by the standard proteasome but not by the immunoproteasome results in poor presentation by dendritic cells. Immunity. 2000;12(1):107-17.
187. Center CHG. GeneCards Summary for SFMBT1 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=SFMBT1.
188. National Center of Biotechnology Information. Entrez Gene Summary for OVOL1 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/5017.
189. Center CHG. GeneCards Summary for OVOL1 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=OVOL1.
190. National Center of Biotechnology Information. Entrez Gene Summary for IGF1R Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/3480.
191. Center CHG. GeneCards Summary for IGF1R Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=IGF1R.
192. National Center of Biotechnology Information. Entrez Gene Summary for VEGFA Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/7422.
193. Center CHG. GeneCards Summary for VEGFA Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=VEGFA.
194. National Center of Biotechnology Information. Entrez Gene Summary for A1CF Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/29974.
195. Center CHG. GeneCards Summary for A1CF Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=A1CF.
196. National Center of Biotechnology Information. Entrez Gene Summary for BAZ1B Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/9031.
197. Center CHG. GeneCards Summary for BAZ1B Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=BAZ1B.
198. National Center of Biotechnology Information. Entrez Gene Summary for UBE2Q2 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/92912.

199. Center CHG. GeneCards Summary for UBE2Q2 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=UBE2Q2.
200. National Center of Biotechnology Information. Entrez Gene Summary for ATXN2 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/6311.
201. Center CHG. GeneCards Summary for ATXN2 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=ATXN2.
202. National Center of Biotechnology Information. Entrez Gene Summary for HLF Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/3131.
203. Center CHG. GeneCards Summary for HLF Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=HLF.
204. National Center of Biotechnology Information. Entrez Gene Summary for BCAS3 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/54828.
205. Center CHG. GeneCards Summary for BCAS3 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=BCAS3.
206. National Center of Biotechnology Information. Entrez Gene Summary for ORC4L Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/5000.
207. Center CHG. GeneCards Summary for ORC4 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=ORC4L.
208. National Center of Biotechnology Information. Entrez Gene Summary for INHBB Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/3625.
209. Center CHG. GeneCards Summary for INHBB Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=INHBB.
210. National Center of Biotechnology Information. Entrez Gene Summary for NFAT5 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/10725.
211. Center CHG. GeneCards Summary for NFAT5 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=NFAT5.
212. National Center of Biotechnology Information. Entrez Gene Summary for PRKAG2 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/51422.
213. Center CHG. GeneCards Summary for PRKAG2 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=PRKAG2.
214. National Center of Biotechnology Information. Entrez Gene Summary for MAF Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/4094.
215. Center CHG. GeneCards Summary for MAF Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=MAF.
216. National Center of Biotechnology Information. Entrez Gene Summary for STC1 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/6781.

217. Center CHG. GeneCards Summary for STC1 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=STC1.
218. National Center of Biotechnology Information. Entrez Gene Summary for PRPSAP1 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/5635.
219. Center CHG. GeneCards Summary for PRPSAP1 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=PRPSAP1.
220. National Center of Biotechnology Information. Entrez Gene Summary for TRIM46 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/80128.
221. Center CHG. GeneCards Summary for TRIM46 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=TRIM46.
222. National Center of Biotechnology Information. Entrez Gene Summary for TMEM171 Gene 2017 [cited 2018 03 January]. Available from: https://www.ncbi.nlm.nih.gov/gene/134285.
223. Center CHG. GeneCards Summary for TMEM171 Gene 2017 [cited 2018 03 January]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=TMEM171.
224. Alvarez-Nemegyei J, Medina-Escobedo M, Villanueva-Jorge S, Vazquez-Mellado J. Prevalence and risk factors for urolithiasis in primary gout: is a reappraisal needed? The Journal of rheumatology. 2005;32(11):2189-91.
225. Borghi C, Rosei EA, Bardin T, Dawson J, Dominiczak A, Kielstein JT, et al. Serum uric acid and the risk of cardiovascular and renal disease. Journal of hypertension. 2015;33(9):1729-41; discussion 41.
226. Nakanishi N, Okamoto M, Yoshida H, Matsuo Y, Suzuki K, Tatara K. Serum uric acid and risk for development of hypertension and impaired fasting glucose or Type II diabetes in Japanese male office workers. European journal of epidemiology. 2003;18(6):523-30.
227. Soltani Z, Rasheed K, Kapusta DR, Reisin E. Potential role of uric acid in metabolic syndrome, hypertension, kidney injury, and cardiovascular diseases: is it time for reappraisal? Current hypertension reports. 2013;15(3):175-81.
228. Emmerson BT, Nagel SL, Duffy DL, Martin NG. Genetic control of the renal clearance of urate: a study of twins. Annals of the rheumatic diseases. 1992;51(3):375-7.
229. Wilk JB, Djousse L, Borecki I, Atwood LD, Hunt SC, Rich SS, et al. Segregation analysis of serum uric acid in the NHLBI Family Heart Study. Human genetics. 2000;106(3):355-9.
230. Dehghan A, Kottgen A, Yang Q, Hwang SJ, Kao WL, Rivadeneira F, et al. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. Lancet. 2008;372(9654):1953-61.
231. Phipps-Green AJ, Merriman ME, Topless R, Altaf S, Montgomery GW, Franklin C, et al. Twenty-eight loci that influence serum urate levels: analysis of association with gout. Annals of the rheumatic diseases. 2014.
232. Tan-Koi WC, Sung C, Chong YY, Lateef A, Pang SM, Vasudevan A, et al. Tailoring of recommendations to reduce serious cutaneous adverse drug reactions: A pharmacogenomics approach. Pharmacogenomics. 2017;18(9):865-74.
233. Garrod AB. Observations on certain pathological conditions of the blood and urine, in gout, rheumatism, and Bright's disease. Medico-chirurgical transactions. 1848;31:83-97.

234. Davis NS, Jr. The cardio-vascular and renal relations and manifestations of gout. Jama. 1897;XXIX(6):261-2.
235. Freedman DS, Williamson DF, Gunter EW, Byers T. Relation of serum uric acid to mortality and ischemic heart disease. The NHANES I Epidemiologic Follow-up Study. American journal of epidemiology. 1995;141(7):637-44.
236. Culleton BF, Larson MG, Kannel WB, Levy D. Serum uric acid and risk for cardiovascular disease and death: the Framingham Heart Study. Annals of internal medicine. 1999;131(1):7-13.
237. Wannamethee SG, Shaper AG, Whincup PH. Serum urate and the risk of major coronary heart disease events. Heart. 1997;78(2):147-53.
238. Vaccarino V, Krumholz HM. Risk factors for cardiovascular disease: one down, many more to evaluate. Annals of internal medicine. 1999;131(1):62-3.
239. Feig DI, Kang D-H, Johnson RJ. Uric Acid and Cardiovascular Risk. The New England journal of medicine. 2008;359(17):1811-21.
240. Nakagawa T, Kang DH, Feig D, Sanchez-Lozada LG, Srinivas TR, Sautin Y, et al. Unearthing uric acid: an ancient factor with recently found significance in renal and cardiovascular disease. Kidney international. 2006;69(10):1722-5.
241. George J, Struthers AD. Role of urate, xanthine oxidase and the effects of allopurinol in vascular oxidative stress. Vascular health and risk management. 2009;5(1):265-72.
242. Meotti FC, Jameson GN, Turner R, Harwood DT, Stockwell S, Rees MD, et al. Urate as a physiological substrate for myeloperoxidase: implications for hyperuricemia and inflammation. The Journal of biological chemistry. 2011;286(15):12901-11.
243. Nakagawa T, Hu H, Zharikov S, Tuttle KR, Short RA, Glushakova O, et al. A causal role for uric acid in fructose-induced metabolic syndrome. American journal of physiology Renal physiology. 2006;290(3):F625-31.
244. Pasalic D, Marinkovic N, Feher-Turkovic L. Uric acid as one of the important factors in multifactorial disorders--facts and controversies. Biochemia medica. 2012;22(1):63-75.
245. Alvarez-Lario B, Macarron-vicente J. Is there anything good in uric acid? Qjm. 2011;104(12):1015-24.
246. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled clinical trials. 1986;7(3):177-88.
247. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society Series A, (Statistics in Society). 2009;172(1):137-59.
248. Higgins JP. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. International journal of epidemiology. 2008;37(5):1158-60.
249. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ (Clinical research ed). 1997;315(7109):629-34.
250. Ioannidis JP. Excess significance bias in the literature on brain volume abnormalities. Archives of general psychiatry. 2011;68(8):773-80.
251. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clinical trials (London, England). 2007;4(3):245-53.
252. Ioannidis JPA. Clarifications on the application and interpretation of the test for excess significance and its extensions. J Math Psychol. 2013;57(5):184-7.

253. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Statistics in medicine. 2000;19(22):3127-31.
254. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. International journal of epidemiology. 2013;42(5):1497-501.
255. Bellou V, Belbasis L, Tzoulaki I, Evangelou E, Ioannidis JP. Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses. Parkinsonism & related disorders. 2016;23:1-9.
256. Kavanagh BP. The GRADE system for rating clinical guidelines. PLoS medicine. 2009;6(9):e1000094.
257. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. International journal of epidemiology. 2011;40(3):740-52.
258. Baker JF, Krishnan E, Chen L, Schumacher HR. Serum uric acid and cardiovascular disease: recent developments, and where do they leave us? The American journal of medicine. 2005;118(8):816-26.
259. Strazzullo P, Puig JG. Uric acid and oxidative stress: relative impact on cardiovascular risk? Nutr Metab Cardiovasc Dis. 2007;17(6):409-14.
260. Barron E, Lara J, White M, Mathers JC. Blood-borne biomarkers of mortality risk: Systematic review of cohort studies. PloS one. 2015;10(6).
261. Dimitroula HV, Hatzitolios AI, Karvounis HI. The role of uric acid in stroke: The issue remains unresolved. Neurologist. 2008;14(4):238-42.
262. Hwu CM, Lin KH. Uric acid and the development of hypertension. Medical Science Monitor. 2010;16(10):RA224-30.
263. Avram Z, Krishnan E. Hyperuricaemia - Where nephrology meets rheumatology. Rheumatology. 2008;47(7):960-4.
264. Feig DI. Uric acid: a novel mediator and marker of risk in chronic kidney disease? Current opinion in nephrology and hypertension. 2009;18(6):526-30.
265. Alonso A, Sovell KA. Gout, hyperuricemia, and Parkinson's disease: a protective effect? Current rheumatology reports. 2010;12(2):149-55.
266. Chang YT, Chang WN, Tsai NW, Huang CC, Kung CT, Su YJ, et al. The roles of biomarkers of oxidative stress and antioxidant in alzheimer's disease: A systematic review. Biomed Res Int. 2014;2014(182303).
267. Cnossen JS, de Ruyter-Hanhijarvi H, van der Post JA, Mol BW, Khan KS, ter Riet G. Accuracy of serum uric acid determination in predicting pre-eclampsia: a systematic review. Acta Obstet Gynecol Scand. 2006;85(5):519-25.
268. Zhang CH, Huang DS, Shen D, Zhang LW, Ma YJ, Wang YM, et al. Association Between Serum Uric Acid Levels and Atrial Fibrillation Risk. Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology. 2016;38(4):1589-95.
269. Zhao J, Liu T, Korantzopoulos P, Letsas KP, Zhang E, Yang Y, et al. Association between serum uric acid and atrial fibrillation recurrence following catheter ablation: A meta-analysis. International journal of cardiology. 2016;204:103-5.
270. Qin T, Zhou X, Wang J, Wu X, Li Y, Wang L, et al. Hyperuricemia and the Prognosis of Hypertensive Patients: A Systematic Review and Meta-Analysis. Journal of clinical hypertension (Greenwich, Conn). 2016.

271. Huang H, Huang B, Li Y, Huang Y, Li J, Yao H, et al. Uric acid and risk of heart failure: A systematic review and meta-analysis. Eur J Heart Fail. 2014;16(1):15-24.
272. Wang J, Qin T, Chen J, Li Y, Wang L, Huang H, et al. Hyperuricemia and risk of incident hypertension: A systematic review and meta-analysis of observational studies. PloS one. 2014;9(12).
273. Jiang M, Gong D, Fan Y. Serum uric acid levels and risk of prehypertension: a meta-analysis. Clinical chemistry and laboratory medicine. 2016.
274. Song X, Hou X, Che K, Wang R, Liu Y, Wang Y, et al. Association between hyperuricemia and clinical adverse outcomes after percutaneous coronary intervention: A meta-analysis. Int J Cardiovasc 2015;201:658-62.
275. Trkulja V, Car S. On-admission serum uric acid predicts outcomes after acute myocardial infarction: systematic review and meta-analysis of prognostic studies. Croat Med J. 2012;53(2):162-72.
276. Li M, Hou W, Zhang X, Hu L, Tang Z. Hyperuricemia and risk of stroke: A systematic review and meta-analysis of prospective studies. Atherosclerosis. 2014;232(2):265-70.
277. Kodama S, Saito K, Yachi Y, Asumi M, Sugawara A, Totsuka K, et al. Association between serum uric acid and development of type 2 diabetes. Diabetes Care. 2009;32(9):1737-42.
278. Jia Z, Zhang X, Kang S, Wu Y. Serum uric acid levels and incidence of impaired fasting glucose and type 2 diabetes mellitus: A meta-analysis of cohort studies. Diabetes Res Clin Pract. 2013;101(1):88-96.
279. Xu Y, Zhu J, Gao L, Liu Y, Shen J, Shen C, et al. Hyperuricemia as an independent predictor of vascular complications and mortality in type 2 diabetes patients: a meta-analysis. PloS one. 2013;8(10):e78206.
280. Xu X, Du N, Wang R, Wang Y, Cai S. Hyperuricemia is independently associated with increased risk of atrial fibrillation: A meta-analysis of cohort studies. International journal of cardiology. 2015;184(1):699-702.
281. Yu S, Chen Y, Hou X, Xu D, Che K, Li C, et al. Serum Uric Acid Levels and Diabetic Peripheral Neuropathy in Type 2 Diabetes: a Systematic Review and Meta-analysis. Molecular neurobiology. 2016;53(2):1045-51.
282. Zhu P, Liu Y, Han L, Xu G, Ran JM. Serum uric acid is associated with incident chronic kidney disease in middle-aged populations: A meta-analysis of 15 cohort studies. PloS one. 2014;9(6).
283. Li L, Yang C, Zhao Y, Zeng X, Liu F, Fu P. Is hyperuricemia an independent risk factor for new-onset chronic kidney disease?: A systematic review and meta-analysis based on observational cohort studies. BMC Nephrology. 2014;15(1).
284. Huang Y, Li YL, Huang H, Wang L, Yuan WM, Li J. Effects of hyperuricemia on renal function of renal transplant recipients: A systematic review and Meta-analysis of cohort studies. PloS one. 2012;7(6).
285. Du N, Xu D, Hou X, Song X, Liu C, Chen Y, et al. Inverse Association Between Serum Uric Acid Levels and Alzheimer's Disease Risk. Molecular neurobiology. 2016;53(4):2594-9.
286. Khan A, T JQ, Dawson J. Serum uric acid level and cognitive function; A systematic review and meta-analysis. Cerebrovasc Dis. 2013;35:821.

287. Shen C, Guo Y, Luo W, Lin C, Ding M. Serum urate and the risk of Parkinson's disease: Results from a meta-analysis. Can J Neurol Sci. 2013;40(1):73-9.
288. Wang L, Hu W, Wang J, Qian W, Xiao H. Low serum uric acid levels in patients with multiple sclerosis and neuromyelists opticis: An updated meta-analysis. Mult Scler Relat Disord. 2016;9:17-22.
289. Abraham A, Drory VE. Influence of serum uric acid levels on prognosis and survival in amyotrophic lateral sclerosis: a meta-analysis. J Neurol. 2014;261(6):1133-8.
290. Flatow J, Buckley P, Miller BJ. Meta-analysis of oxidative stress in schizophrenia. Biol Psychiatry. 2013;74(6):400-9.
291. Bartoli F, Crocamo C, Mazza MG, Clerici M, Carrà G. Uric acid levels in subjects with bipolar disorder: A comparative meta-analysis. J Psychiatr Res. 2016.
292. Yan S, Zhang P, Xu W, Liu Y, Wang B, Jiang T, et al. Serum Uric Acid Increases Risk of Cancer Incidence and Mortality: A Systematic Review and Meta-Analysis. Mediators Inflamm. 2015;2015(764250).
293. Zhao G. Baseline serum uric acid level as a predictor of cardiovascular disease related mortality and all-cause mortality: A meta-analysis of prospective studies. Cardiology (Switzerland). 2014;129:47.
294. Xia X, Luo Q, Li B, Lin Z, Yu X, Huang F. Serum uric acid and mortality in chronic kidney disease: a systematic review and meta-analysis. Metabolism: clinical and experimental. 2016;65(9):1326-41.
295. Tamariz L, Harzand A, Verma S, Jones J, Hare J. Uric acid as a predictor of mortality in congestive heart failure: A meta-analysis. J Am Coll Cardiol. 2009;53 (10):A169.
296. Yan L, Liu Z, Zhang C. Uric acid as a predictor of in-hospital mortality in acute myocardial infarction: a meta-analysis. Cell Biochem Biophys. 2014;70(3):1597-601.
297. Wang Z, Lin Y, Liu Y, Chen Y, Wang B, Li C, et al. Serum Uric Acid Levels and Outcomes After Acute Ischemic Stroke. Molecular neurobiology. 2016;53(3):1753-9.
298. Li X, Miao X, Wang H, Wang Y, Li F, Yang Q, et al. Association of Serum Uric Acid Levels in Psoriasis: A Systematic Review and Meta-Analysis. Medicine. 2016;95(19):e3676.
299. Zhou Y, Wei F, Fan Y. High serum uric acid and risk of nonalcoholic fatty liver disease: A systematic review and meta-analysis. Clinical biochemistry. 2016;49(7-8):636-42.
300. Braga F, Pasqualetti S, Ferraro S, Panteghini M. Hyperuricemia (HUA) as an independent risk factor for coronary heart disease (CHD) in the general population: A systematic review and meta-analysis. Clin Chem Lab Med. 2015;53:S1199.
301. Chen X, Guo X, Huang R, Chen Y, Zheng Z, Shang H. Serum uric acid levels in patients with Alzheimer's disease: A meta-analysis. PloS one. 2014;9(4).
302. Grayson PC, Kim SY, LaValley M, Choi HK. Hyperuricemia and incident hypertension: a systematic review and meta-analysis. Arthritis Care Res 2011;63(1):102-10.
303. Kim SY, Guevara JP, Kim KM, Choi HK, Heitjan DF, Albert DA. Hyperuricemia and risk of stroke: a systematic review and meta-analysis. Arthritis and rheumatism. 2009;61(7):885-92.
304. Kim SY, Guevara JP, Kim KM, Choi HK, Heitjan DF, Albert DA. Hyperuricemia and coronary heart disease: a systematic review and meta-analysis. . Arthritis Care Res [Internet]. 2010; 62(2):[170-80 pp.].

305. Li YL, Wang L, Li J, Huang Y, Yuan WM. [The correlation between uric acid and the incidence and prognosis of kidney diseases: a systematic review and meta-analysis of cohort studies]. [Chinese]. Zhonghua nei ke za zhi [Chinese journal of internal medicine]. 2011;50(7):555-61.
306. Liu B, Shen Y, Xiao K, Tang Y, Cen L, Wei J. Serum uric acid levels in patients with multiple sclerosis: A meta-analysis. Neurol Res 2012;34(2):163-71.
307. Lv Q, Meng XF, He FF, Chen S, Su H, Xiong J, et al. High Serum Uric Acid and Increased Risk of Type 2 Diabetes: A Systemic Review and Meta-Analysis of Prospective Cohort Studies. PloS one. 2013;8(2).
308. Schrag M, Mueller C, Zabel M, Crofton A, Kirsch WM, Ghribi O, et al. Oxidative stress in blood in Alzheimer's disease and mild cognitive impairment: A meta-analysis. Neurobiol Dis. 2013;59:100-10.
309. Shen L, Ji HF. Low uric acid levels in patients with Parkinson's disease: Evidence from meta-analysis. BMJ Open. 2013;3(11).
310. Tamariz L, Hernandez F, Bush A, Palacio A, Hare JM. Association between serum uric acid and atrial fibrillation: A systematic review and meta-analysis. Heart Rhythm. 2014;11(7):1102-8.
311. Wheeler JG, Juzwishin KDM, Eiriksdottir G, Gudnason V, Danesh J. Serum uric acid and coronary heart disease in 9,458 incident cases and 155,084 controls: Prospective study and meta-analysis. PLoS medicine. 2005;2(3):0236-43.
312. Yang Y, Fan Y, Li J, Liu B, Shao Y, Cao H, et al. Serum uric acid as a predictor for cardiovascular and all-cause mortality in women versus men. International journal of cardiology. 2015;185:125-8.
313. Zhang W, Sun K, Yang Y, Zhang H, Hu FB, Hui R. Plasma uric acid and hypertension in a Chinese community: prospective study and metaanalysis. Clin Chem. 2009;55(11):2026-34.
314. Fink HA, Wilt TJ, Eidman KE, Garimella PS, MacDonald R, Rutks IR, et al. Medical management to prevent recurrent nephrolithiasis in adults: A systematic review for an American College of Physicians Clinical Guideline. Annals of internal medicine. 2013;158(7):535-43.
315. Wang H, Wei Y, Kong X, Xu D. Effects of urate-lowering therapy in hyperuricemia on slowing the progression of renal function: A meta-analysis. J Ren Nutr. 2013;23(5):389-96.
316. Zhang YF, He F, Ding HH, Dai W, Zhang Q, Luan H, et al. Effect of uric-acid-lowering therapy on progression of chronic kidney disease: a meta-analysis. J Huazhong Univ Sci Technolog Med Sci. 2014;34(4):476-81.
317. Bose B, Badve SV, Hiremath SS, Boudville N, Brown FG, Cass A, et al. Effects of uric acid-lowering therapy on renal outcomes: A systematic review and meta-analysis. Nephrol Dial Transplant 2014;29(2):406-13.
318. Higgins P, Dawson J, Lees KR, McArthur K, Quinn TJ, Walters MR. Xanthine oxidase inhibition for the treatment of cerebrovascular and cardiovascular disease: A systematic review and meta-analysis. Cerebrovasc Dis. 2010;29:321.
319. Chaudhari T, McGuire W. Allopurinol for preventing mortality and morbidity in newborn infants with hypoxic-ischaemic encephalopathy. Cochrane Database Syst Rev [Internet]. 2012; (7).

320. Agarwal V, Hans N, Messerli FH. Effect of Allopurinol on Blood Pressure: A Systematic Review and Meta-Analysis. Journal of clinical hypertension (Greenwich, Conn). 2013;15(6):435-42.
321. Korostishevsky M, Steves CJ, Malkin I, Spector T, Williams FM, Livshits G. Genomics and metabolomics of muscular mass in a community-based sample of UK females. European journal of human genetics : EJHG. 2016;24(2):277-83.
322. Lyngdoh T, Vuistiner P, Marques-Vidal P, Rousson V, Waeber G, Vollenweider P, et al. Serum uric acid and adiposity: deciphering causality using a bidirectional Mendelian randomization approach. PloS one. 2012;7(6):e39321.
323. White J, Sofat R, Hemani G, Shah T, Engmann J, Dale C, et al. Plasma urate concentration and risk of coronary heart disease: a Mendelian randomisation analysis. The lancet Diabetes & endocrinology. 2016;4(4):327-36.
324. Dalbeth N, Topless R, Flynn T, Cadzow M, Bolland MJ, Merriman TR. Mendelian randomization analysis to examine for a causal effect of urate on bone mineral density. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2015;30(6):985-91.
325. Xiong A, Yao Q, He J, Fu W, Yu J, Zhang Z. No causal effect of serum urate on bone-related outcomes among a population of postmenopausal women and elderly men of Chinese Han ethnicity--a Mendelian randomization study. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2016;27(3):1031-9.
326. Kleber ME, Delgado G, Grammer TB, Silbernagel G, Huang J, Kramer BK, et al. Uric Acid and Cardiovascular Events: A Mendelian Randomization Study. Journal of the American Society of Nephrology : JASN. 2015;26(11):2831-8.
327. Keenan T, Zhao W, Rasheed A, Ho WK, Malik R, Felix JF, et al. Causal Assessment of Serum Urate Levels in Cardiometabolic Diseases Through a Mendelian Randomization Study. J Am Coll Cardiol. 2016;67(4):407-16.
328. Palmer TM, Nordestgaard BG, Benn M, Tybjaerg-Hansen A, Davey Smith G, Lawlor DA, et al. Association of plasma uric acid with ischaemic heart disease and blood pressure: mendelian randomisation analysis of two large cohorts. BMJ (Clinical research ed). 2013;347:f4262.
329. Oikonen M, Wendelin-Saarenhovi M, Lyytikainen LP, Siitonen N, Loo BM, Jula A, et al. Associations between serum uric acid and markers of subclinical atherosclerosis in young adults. The cardiovascular risk in Young Finns study. Atherosclerosis. 2012;223(2):497-503.
330. Yan D, Wang J, Jiang F, Zhang R, Wang T, Wang S, et al. A causal relationship between uric acid and diabetic macrovascular disease in Chinese type 2 diabetes patients: A Mendelian randomization analysis. International journal of cardiology. 2016;214:194-9.
331. Mallamaci F, Testa A, Leonardis D, Tripepi R, Pisano A, Spoto B, et al. A genetic marker of uric acid level, carotid atherosclerosis, and arterial stiffness: a family-based study. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2015;65(2):294-302.
332. Sluijs I, Holmes MV, van der Schouw YT, Beulens JW, Asselbergs FW, Huerta JM, et al. A Mendelian Randomization Study of Circulating Uric Acid and Type 2 Diabetes. Diabetes. 2015;64(8):3028-36.

333. Yang Q, Kottgen A, Dehghan A, Smith AV, Glazer NL, Chen MH, et al. Multiple genetic loci influence serum urate levels and their relationship with gout and cardiovascular disease risk factors. Circulation Cardiovascular genetics. 2010;3(6):523-30.
334. Dai X, Yuan J, Yao P, Yang B, Gui L, Zhang X, et al. Association between serum uric acid and the metabolic syndrome among a middle- and old-age Chinese population. European journal of epidemiology. 2013;28(8):669-76.
335. Greenberg KI, McAdams-DeMarco MA, Kottgen A, Appel LJ, Coresh J, Grams ME. Plasma Urate and Risk of a Hospital Stay with AKI: The Atherosclerosis Risk in Communities Study. Clinical journal of the American Society of Nephrology : CJASN. 2015;10(5):776-83.
336. Testa A, Mallamaci F, Spoto B, Pisano A, Sanguedolce MC, Tripepi G, et al. Association of a polymorphism in a gene encoding a urate transporter with CKD progression. Clinical journal of the American Society of Nephrology : CJASN. 2014;9(6):1059-65.
337. Hughes K, Flynn T, de Zoysa J, Dalbeth N, Merriman TR. Mendelian randomization analysis associates increased serum urate, due to genetic variation in uric acid transporters, with improved renal function. Kidney international. 2014;85(2):344-51.
338. Gao J, Xu H, Huang X, Chen H. Short communication: genetic variations of SLC2A9 in relation to Parkinson's disease. Translational neurodegeneration. 2013;2(1):5.
339. Facheris MF, Hicks AA, Minelli C, Hagenah JM, Kostic V, Campbell S, et al. Variation in the uric acid transporter gene SLC2A9 and its association with AAO of Parkinson's disease. Journal of molecular neuroscience : MN. 2011;43(3):246-50.
340. Lyngdoh T, Bochud M, Glaus J, Castelao E, Waeber G, Vollenweider P, et al. Associations of serum uric acid and SLC2A9 variant with depressive and anxiety disorders: a population-based study. PloS one. 2013;8(10):e76336.
341. Houlihan LM, Wyatt ND, Harris SE, Hayward C, Gow AJ, Marioni RE, et al. Variation in the uric acid transporter gene (SLC2A9) and memory performance. Human molecular genetics. 2010;19(11):2321-30.
342. Burgess S, Daniel RM, Butterworth AS, Thompson SG. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. International journal of epidemiology. 2015;44(2):484-95.
343. Gonzalez-Aramburu I, Sanchez-Juan P, Jesus S, Gorostidi A, Fernandez-Juan E, Carrillo F, et al. Genetic variability related to serum uric acid concentration and risk of Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society. 2013;28(12):1737-40.
344. Gonzalez-Aramburu I, Sanchez-Juan P, Sierra M, Fernandez-Juan E, Sanchez-Quintana C, Berciano J, et al. Serum uric acid and risk of dementia in Parkinson's disease. Parkinsonism & related disorders. 2014;20(6):637-9.
345. Han X, Gui L, Liu B, Wang J, Li Y, Dai X, et al. Associations of the uric acid related genetic variants in SLC2A9 and ABCG2 loci with coronary heart disease risk. BMC genetics. 2015;16:4.
346. Mallamaci F, Testa A, Leonardis D, Tripepi R, Pisano A, Spoto B, et al. A polymorphism in the major gene regulating serum uric acid associates with clinic SBP and the white-coat effect in a family-based study. Journal of hypertension. 2014;32(8):1621-8; discussion 8.
347. McKeigue PM, Campbell H, Wild S, Vitart V, Hayward C, Rudan I, et al. Bayesian methods for instrumental variable analysis with genetic instruments ('Mendelian

randomization'): example with urate transporter SLC2A9 as an instrumental variable for effect of urate levels on metabolic syndrome. International journal of epidemiology. 2010;39(3):907-18.
348. Parsa A, Brown E, Weir MR, Fink JC, Shuldiner AR, Mitchell BD, et al. Genotype-based changes in serum uric acid affect blood pressure. Kidney international. 2012;81(5):502-7.
349. Pfister R, Barnes D, Luben R, Forouhi NG, Bochud M, Khaw KT, et al. No evidence for a causal link between uric acid and type 2 diabetes: A Mendelian randomisation approach. Diabetologia. 2011;54(10):2561-9.
350. Rasheed H, Hughes K, Flynn TJ, Merriman TR. Mendelian randomization provides no evidence for a causal role of serum urate in increasing serum triglyceride levels. Circulation Cardiovascular genetics. 2014;7(6):830-7.
351. Sedaghat S, Pazoki R, Uitterlinden AG, Hofman A, Stricker BH, Ikram MA, et al. Association of uric acid genetic risk score with blood pressure: the Rotterdam study. Hypertension (Dallas, Tex : 1979). 2014;64(5):1061-6.
352. Simon KC, Eberly S, Gao X, Oakes D, Tanner CM, Shoulson I, et al. Mendelian randomization of serum urate and parkinson disease progression. Annals of neurology. 2014;76(6):862-8.
353. Stark K, Reinhard W, Grassl M, Erdmann J, Schunkert H, Illig T, et al. Common polymorphisms influencing serum uric acid levels contribute to susceptibility to gout, but not to coronary artery disease. PloS one. 2009;4(11):e7729.
354. Sun X, Zhang R, Jiang F, Tang S, Chen M, Peng D, et al. Common variants related to serum uric acid concentrations are associated with glucose metabolism and insulin secretion in a Chinese population. PloS one. 2015;10(1):e0116714.
355. Tabara Y, Kohara K, Kawamoto R, Hiura Y, Nishimura K, Morisaki T, et al. Association of four genetic loci with uric acid levels and reduced renal function: the J-SHIPP Suita study. American journal of nephrology. 2010;32(3):279-86.
356. Voruganti VS, Franceschini N, Haack K, Laston S, MacCluer JW, Umans JG, et al. Replication of the effect of SLC2A9 genetic variation on serum uric acid levels in American Indians. European journal of human genetics : EJHG. 2014;22(7):938-43.
357. Holme I, Aastveit AH, Hammar N, Jungner I, Walldius G. Uric acid and risk of myocardial infarction, stroke and congestive heart failure in 417,734 men and women in the Apolipoprotein MOrtality RISk study (AMORIS). Journal of internal medicine. 2009;266(6):558-70.
358. Mazza A, Pessina AC, Pavei A, Scarpa R, Tikhonoff V, Casiglia E. Predictors of stroke mortality in elderly people from the general population. The CArdiovascular STudy in the ELderly. European journal of epidemiology. 2001;17(12):1097-104.
359. Kuo CF, See LC, Yu KH, Chou IJ, Chiou MJ, Luo SF. Significance of serum uric acid levels on the risk of all-cause and cardiovascular mortality. Rheumatology (Oxford, England). 2013;52(1):127-34.
360. Chen JH, Chuang SY, Chen HJ, Yeh WT, Pan WH. Serum uric acid level as an independent risk factor for all-cause, cardiovascular, and ischemic stroke mortality: a Chinese cohort study. Arthritis and rheumatism. 2009;61(2):225-32.
361. Owens DK, Lohr KN, Atkins D, Treadwell JR, Reston JT, Bass EB, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical

interventions--agency for healthcare research and quality and the effective health-care program. Journal of clinical epidemiology. 2010;63(5):513-23.
362. Theodoratou E, Tzoulaki I, Zgaga L, Ioannidis JP. Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials. BMJ (Clinical research ed). 2014;348:g2035.
363. Belbasis L, Savvidou MD, Kanu C, Evangelou E, Tzoulaki I. Birth weight in relation to health and disease in later life: an umbrella review of systematic reviews and meta-analyses. BMC medicine. 2016;14(1):147.
364. Krishnan E, Lessov-Schlaggar CN, Krasnow RE, Swan GE. Nature versus nurture in gout: a twin study. The American journal of medicine. 2012;125(5):499-504.
365. Chen H, Mosley TH, Alonso A, Huang X. Plasma urate and Parkinson's disease in the Atherosclerosis Risk in Communities (ARIC) study. American journal of epidemiology. 2009;169(9):1064-9.
366. Weisskopf MG, O'Reilly E, Chen H, Schwarzschild MA, Ascherio A. Plasma urate and risk of Parkinson's disease. American journal of epidemiology. 2007;166(5):561-7.
367. Kim TS, Pae CU, Yoon SJ, Jang WY, Lee NJ, Kim JJ, et al. Decreased plasma antioxidants in patients with Alzheimer's disease. International journal of geriatric psychiatry. 2006;21(4):344-8.
368. Feig DI, Madero M, Jalal DI, Sanchez-Lozada LG, Johnson RJ. Uric acid and the origins of hypertension. The Journal of pediatrics. 2013;162(5):896-902.
369. Tamariz L, Hare JM. Xanthine oxidase inhibitors in heart failure: where do we go from here? Circulation. 2015;131(20):1741-4.
370. Ioannidis JP. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. The Milbank quarterly. 2016;94(3):485-514.
371. Kavvoura FK, Liberopoulos G, Ioannidis JP. Selection in reported epidemiological risks: an empirical assessment. PLoS medicine. 2007;4(3):e79.
372. Ioannidis JP. Exposure-wide epidemiology: revisiting Bradford Hill. Statistics in medicine. 2016;35(11):1749-62.
373. Baker JF, Krishnan E, Chen L, Schumacher HR. Serum uric acid and cardiovascular disease: recent developments, and where do they leave us? American Journal of Medicine. 2005;118(8):816-26.
374. Strazzullo P, Puig JG. Uric acid and oxidative stress: relative impact on cardiovascular risk? Nutrition Metabolism & Cardiovascular Diseases. 2007;17(6):409-14.
375. Feig DI. Uric acid: a novel mediator and marker of risk in chronic kidney disease? Current Opinion in Nephrology & Hypertension. 2009;18(6):526-30.
376. Alonso A, Sovell KA. Gout, hyperuricemia, and Parkinson's disease: A protective effect? Current Rheumatology Reports. 2010;12(2):149-55.
377. Chang YT, Chang WN, Tsai NW, Huang CC, Kung CT, Su YJ, et al. The roles of biomarkers of oxidative stress and antioxidant in alzheimer's disease: A systematic review. BioMed Research International. 2014;2014(182303).
378. Cnossen JS, de Ruyter-Hanhijarvi H, van der Post JA, Mol BW, Khan KS, ter Riet G. Accuracy of serum uric acid determination in predicting pre-eclampsia: a systematic review. Acta Obstetricia et Gynecologica Scandinavica. 2006;85(5):519-25.
379. Xu X, Du N, Wang R, Wang Y, Cai S. Hyperuricemia is independently associated with increased risk of atrial fibrillation: A meta-analysis of cohort studies. International Journal of Cardiology. 2015;184(1):699-702.

380. Wheeler JG, Juzwishin KDM, Eiriksdottir G, Gudnason V, Danesh J. Serum uric acid and coronary heart disease in 9,458 incident cases and 155,084 controls: Prospective study and meta-analysis. PLoS Medicine. 2005;2(3):0236-43.
381. Kim SY, Guevara JP, Kim KM, Choi HK, Heitjan DF, Albert DA. Hyperuricemia and coronary heart disease: a systematic review and meta-analysis (Provisional abstract). Arthritis Care and Research [Internet]. 2010; 62(2):[170-80 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/cldare/articles/DARE-12010002709/frame.html

http://onlinelibrary.wiley.com/store/10.1002/acr.20065/asset/20065_ftp.pdf?v=1&t=igy1oupl&s=fbd5e9cab727198004e2c60c8a8b074c66db0729

http://onlinelibrary.wiley.com/store/10.1002/acr.20065/asset/20065_ftp.pdf?v=1&t=ija4znix&s=91c42016aa8497c31f32643d48e8e0234f673b3e.
382. Braga F, Pasqualetti S, Ferraro S, Panteghini M. Hyperuricemia (HUA) as an independent risk factor for coronary heart disease (CHD) in the general population: A systematic review and meta-analysis. Clinical Chemistry and Laboratory Medicine. 2015;53:S1199.
383. Huang H, Huang B, Li Y, Huang Y, Li J, Yao H, et al. Uric acid and risk of heart failure: A systematic review and meta-analysis. European Journal of Heart Failure. 2014;16(1):15-24.
384. Zhang W, Sun K, Yang Y, Zhang H, Hu FB, Hui R. Plasma uric acid and hypertension in a Chinese community: Prospective study and metaanalysis. Clinical Chemistry. 2009;55(11):2026-34.
385. Grayson PC, Kim SY, LaValley M, Choi HK. Hyperuricemia and incident hypertension: a systematic review and meta-analysis. Arthritis care & research. 2011;63(1):102-10.
386. Yan L, Liu Z, Zhang C. Uric acid as a predictor of in-hospital mortality in acute myocardial infarction: a meta-analysis. Cell biochemistry and biophysics. 2014;70(3):1597-601.
387. Song X, Hou X, Che K, Wang R, Liu Y, Wang Y, et al. Association between hyperuricemia and clinical adverse outcomes after percutaneous coronary intervention: A meta-analysis. International Journal of Cardiology. 2015;201:658-62.
388. Trkulja V, Car S. On-admission serum uric acid predicts outcomes after acute myocardial infarction: systematic review and meta-analysis of prognostic studies. Croatian Medical Journal. 2012;53(2):162-72.
389. Kim SY, Guevara JP, Kim KM, Choi HK, Heitjan DF, Albert DA. Hyperuricemia and risk of stroke: a systematic review and meta-analysis. Arthritis & Rheumatism. 2009;61(7):885-92.
390. Jia Z, Zhang X, Kang S, Wu Y. Serum uric acid levels and incidence of impaired fasting glucose and type 2 diabetes mellitus: A meta-analysis of cohort studies. Diabetes Research and Clinical Practice. 2013;101(1):88-96.
391. Khan A, T JQ, Dawson J. Serum uric acid level and cognitive function; A systematic review and meta-analysis. Cerebrovascular Diseases. 2013;35:821.
392. Schrag M, Mueller C, Zabel M, Crofton A, Kirsch WM, Ghribi O, et al. Oxidative stress in blood in Alzheimer's disease and mild cognitive impairment: A meta-analysis. Neurobiology of Disease. 2013;59:100-10.

393. Du N, Xu D, Hou X, Song X, Liu C, Chen Y, et al. Inverse Association Between Serum Uric Acid Levels and Alzheimer's Disease Risk. Molecular Neurobiology. 2016;53(4):2594-9.
394. Bartoli F, Crocamo C, Mazza MG, Clerici M, Carrà G. Uric acid levels in subjects with bipolar disorder: A comparative meta-analysis. Journal of Psychiatric Research. 2016.
395. Shen C, Guo Y, Luo W, Lin C, Ding M. Serum urate and the risk of Parkinson's disease: Results from a meta-analysis. Canadian Journal of Neurological Sciences. 2013;40(1):73-9.
396. Liu B, Shen Y, Xiao K, Tang Y, Cen L, Wei J. Serum uric acid levels in patients with multiple sclerosis: A meta-analysis. Neurological Research. 2012;34(2):163-71.
397. Wang L, Hu W, Wang J, Qian W, Xiao H. Low serum uric acid levels in patients with multiple sclerosis and neuromyelists opticis: An updated meta-analysis. Multiple Sclerosis and Related Disorders. 2016;9:17-22.
398. Abraham A, Drory VE. Influence of serum uric acid levels on prognosis and survival in amyotrophic lateral sclerosis: a meta-analysis. Journal of neurology. 2014;261(6):1133-8.
399. Flatow J, Buckley P, Miller BJ. Meta-analysis of oxidative stress in schizophrenia. Biological Psychiatry. 2013;74(6):400-9.
400. Yan S, Zhang P, Xu W, Liu Y, Wang B, Jiang T, et al. Serum Uric Acid Increases Risk of Cancer Incidence and Mortality: A Systematic Review and Meta-Analysis. Mediators of Inflammation. 2015;2015 (no pagination)(764250).
401. Yang Y, Fan Y, Li J, Liu B, Shao Y, Cao H, et al. Serum uric acid as a predictor for cardiovascular and all-cause mortality in women versus men. International Journal of Cardiology. 2015;185:125-8.
402. Tamariz L, Harzand A, Verma S, Jones J, Hare J. Uric acid as a predictor of mortality in congestive heart failure: A meta-analysis. Journal of the American College of Cardiology. 2009;53 (10):A169.
403. Fink HA, Wilt TJ, Eidman KE, Garimella PS, MacDonald R, Rutks IR, et al. Medical management to prevent recurrent nephrolithiasis in adults: A systematic review for an American College of Physicians Clinical Guideline. Annals of Internal Medicine. 2013;158(7):535-43.
404. Wang H, Wei Y, Kong X, Xu D. Effects of urate-lowering therapy in hyperuricemia on slowing the progression of renal function: A meta-analysis. Journal of Renal Nutrition. 2013;23(5):389-96.
405. Zhang YF, He F, Ding HH, Dai W, Zhang Q, Luan H, et al. Effect of uric-acid-lowering therapy on progression of chronic kidney disease: a meta-analysis. Journal of Huazhong University of Science and Technology Medical Sciences. 2014;34(4):476-81.
406. Bose B, Badve SV, Hiremath SS, Boudville N, Brown FG, Cass A, et al. Effects of uric acid-lowering therapy on renal outcomes: A systematic review and meta-analysis. Nephrology Dialysis Transplantation. 2014;29(2):406-13.
407. Higgins P, Dawson J, Lees KR, McArthur K, Quinn TJ, Walters MR. Xanthine oxidase inhibition for the treatment of cerebrovascular and cardiovascular disease: A systematic review and meta-analysis. Cerebrovascular Diseases. 2010;29:321.
408. Kanbay M, Siriopol D, Nistor I, Elcioglu OC, Telci O, Takir M, et al. Effects of allopurinol on endothelial dysfunction: A meta-analysis. American Journal of Nephrology. 2014;39(4):348-56.

409. Chaudhari T, McGuire W. Allopurinol for preventing mortality and morbidity in newborn infants with hypoxic-ischaemic encephalopathy. Cochrane Database of Systematic Reviews [Internet]. 2012; (7). Available from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD006817.pub3/abstract

http://onlinelibrary.wiley.com/store/10.1002/14651858.CD006817.pub3/asset/CD006817.pdf?v=1&t=igy1jpk0&s=4303bbe82e7c797368c533d666af195dc7e82e52.

410. Agarwal V, Hans N, Messerli FH. Effect of Allopurinol on Blood Pressure: A Systematic Review and Meta-Analysis. Journal of Clinical Hypertension. 2013;15(6):435-42.

411. Biobank U. UK Biobank Ethics and Governance Framework 2007 [cited 2017 02 February]. Available from: http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf?phpMyAdmin=trmKQlYdjjnQIgJ%2CfAzikMhEnx6.

412. UK Biobank. Access Procedures: Application and review procedures for access to the UK Biobank Resource 2011 [cited 2018 20 October]. Available from: www.ukbiobank.ac.uk/resources/.

413. Office of Public Sector Information. Data Protection Act 1998 1998 [cited 2015 06 October]. Available from: http://www.legislation.gov.uk/ukpga/1998/29.

414. Biobank U. UK Biobank: Protocol for a large-scale prospective epidemiological resource 2007 [cited 2018 20 January].

415. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine. 2015;12(3):e1001779.

416. Affymetrix Research Services Laboratory. UKB_WCSGAX: UK Biobank 500K Samples Processing 2015 [cited 2018 06 January]. Available from: http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155583.

417. Biobank U. Genotyping of 500,000 UK Biobank participants 2015 [cited 2017 06 January]. Available from: https://biobank.ctsu.ox.ac.uk/crystal/docs/genotyping_sample_workflow.pdf.

418. Affymetrix Research Services Laboratory. UKB_WCSGAX: UK Biobank 500K Samples Genotyping Data Generation 2015 [cited 2018 06 January]. Available from: http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155582.

419. UK Biobank. UK Biobank Axiom Array [cited 2018 06 January]. Available from: http://www.ukbiobank.ac.uk/scientists-3/uk-biobank-axiomarray/.

420. Wellcome Trust Centre for Human Genetics. Genotyping and quality control of UK Biobank, a large-scale, extensively phenotyped, prospective resource 2015 [cited 2018 06 January]. Available from: http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580.

421. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. Genome-wide genetic data on~ 500,000 UK Biobank participants. bioRxiv. 2017.

422. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nature genetics. 2012;44(8):955-9.

423. O'Connell J, Sharp K, Shrine N, Wain L, Hall I, Tobin M, et al. Haplotype estimation for biobank-scale data sets. Nature genetics. 2016;48(7):817-20.

424. Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, et al. The UK10K project identifies rare variants in health and disease. Nature. 2015;526(7571):82-90.

425. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73.
426. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature genetics. 2016;48(10):1279-83.
427. Dilthey A, Leslie S, Moutsianas L, Shen J, Cox C, Nelson MR, et al. Multi-population classical HLA type imputation. PLoS computational biology. 2013;9(2):e1002877.
428. UK Biobank. Integrating electronic health records into the UK Biobank Resource 2014 [cited 2018 07 January]. Available from: https://biobank.ctsu.ox.ac.uk/crystal/docs/DataLinkageProcess.pdf.
429. Biobank U. Mapping inpatient hospital data across England, Scotland and Wales 2014 [cited 2018 07 January]. Available from: http://biobank.ctsu.ox.ac.uk/crystal/docs/inpatient_mapping.pdf.
430. UK Biobank. Cancer data: linkage from national cancer registries 2013 [cited 2017 20 July]. Available from: http://www.ukbiobank.ac.uk/resources/.
431. UK Biobank. Mortality data: linkage from national death registries 2013 [cited 2017 20 July]. Available from: http://www.ukbiobank.ac.uk/resources/.
432. Galinsky KJ, Bhatia G, Loh PR, Georgiev S, Mukherjee S, Patterson NJ, et al. Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. American journal of human genetics. 2016;98(3):456-72.
433. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nature biotechnology. 2013;31(12):1102-10.
434. Ritchie MD, Denny JC, Zuvich RL, Crawford DC, Schildcrout JS, Bastarache L, et al. Genome- and phenome-wide analyses of cardiac conduction identifies markers of arrhythmia risk. Circulation. 2013;127(13):1377-85.
435. Cronin RM, Field JR, Bradford Y, Shaffer CM, Carroll RJ, Mosley JD, et al. Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. Frontiers in genetics. 2014;5:250.
436. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics (Oxford, England). 2010;26(9):1205-10.
437. Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, Bradford Y, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: Using electronic medical records for genome- and phenome-wide studies. American journal of human genetics. 2011;89(4):529-42.
438. Pathak J, Kiefer RC, Bielinski SJ, Chute CG. Mining the human phenome using semantic web technologies: a case study for Type 2 Diabetes. Amia 2012;Annual Symposium proceedings / AMIA Symposium. AMIA Symposium. 2012:699-708.
439. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nature Biotechnology. 2013;31(12):1102-10.
440. Hebbring SJ, Schrodi SJ, Ye Z, Zhou Z, Page D, Brilliant MH. A PheWAS approach in studying HLA-DRB1 1501. Genes Immun. 2013;14(3):187-91.

441. Pendergrass SA, Brown-Gentry K, Dudek S, Frase A, Torstenson ES, Goodloe R, et al. Phenome-Wide Association Study (PheWAS) for Detection of Pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. PLoS genetics. 2013;9(1).
442. Cronin RM, Field JR, Bradford Y, Shaffer CM, Carroll RJ, Mosley JD, et al. Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. Frontiers in Genetics. 2014;5(AUG).
443. Hall MA, Verma A, Brown-Gentry KD, Goodloe R, Boston J, Wilson S, et al. Detection of Pleiotropy through a Phenome-Wide Association Study (PheWAS) of Epidemiologic Data as Part of the Environmental Architecture for Genes Linked to Environment (EAGLE) Study. PLoS Genetics. 2014;10(12).
444. Mitchell SL, Hall JB, Goodloe RJ, Boston J, Farber-Eger E, Pendergrass SA, et al. Investigating the relationship between mitochondrial genetic variation and cardiovascular-related traits to develop a framework for mitochondrial phenome-wide association studies. BioData Mining. 2014;7(1).
445. Namjou B, Marsolo K, Caroll RJ, Denny JC, Ritchie MD, Verma SS, et al. Phenome-wide association study (PheWAS) in EMR-linked pediatric cohorts, genetically links PLCL1 to speech language development and IL5-IL13 to Eosinophilic Esophagitis. Frontiers in genetics. 2014;5(NOV).
446. Shameer K, Denny JC, Ding K, Jouni H, Crosslin DR, De Andrade M, et al. A genome- and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. Human genetics. 2014;133(1):95-109.
447. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. Bioinformatics (Oxford, England). 2014;30(16):2375-6.
448. Ye Z, Mayer J, Ivacic L, Zhou Z, He M, Schrodi SJ, et al. Phenome-wide association studies (PheWASs) for functional variants. European Journal of Human Genetics. 2015;23(4):523-9.
449. Diogo D, Bastarache L, Liao KP, Graham RR, Fulton RS, Greenberg JD, et al. TYK2 protein-coding variants protect against rheumatoid arthritis and autoimmunity, with no evidence of major pleiotropic effects on non-autoimmune complex traits. PloS one. 2015;10(4).
450. Moore CB, Verma A, Pendergrass S, Verma SS, Johnson DH, Daar ES, et al. Phenome-wide association study relating pretreatment laboratory parameters with human genetic variants in AIDS clinical trials group protocols. Open Forum Infectious Diseases. 2015;2(1).
451. Hebbring SJ, Rastegar-Mojarad M, Ye Z, Mayer J, Jacobson C, Lin S. Application of clinical text data for phenome-wide association studies (PheWASs). Bioinformatics. 2015;31(12):1981-7.
452. Warner JL, Alterovitz G. Phenome based analysis as a means for discovering context dependent clinical reference ranges. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2012;2012:1441-9.
453. Boland MR, Hripcsak G, Albers DJ, Wei Y, Wilcox AB, Wei J, et al. Discovering medical conditions associated with periodontitis using linked electronic health records. Journal of clinical periodontology. 2013;40(5):474-82.

454. Liao KP, Cai T, Gainer V, Goryachev S, Zeng-treitler Q, Raychaudhuri S, et al. Electronic medical records for discovery research in rheumatoid arthritis. Arthritis care & research. 2010;62(8):1120-7.
455. Neuraz A, Chouchana L, Malamut G, Le Beller C, Roche D, Beaune P, et al. Phenome-wide association studies on a quantitative trait: application to TPMT enzyme activity and thiopurine therapy in pharmacogenomics. PLoS computational biology. 2013;9(12):e1003405.
456. Warner JL, Zollanvari A, Ding Q, Zhang P, Snyder GM, Alterovitz G. Temporal phenome analysis of a large electronic health record cohort enables identification of hospital-acquired complications. Journal of the American Medical Informatics Association : JAMIA. 2013a;20(e2):e281-7.
457. Warner JL, Alterovitz G, Bodio K, Joyce RM. External phenome analysis enables a rational federated query strategy to detect changing rates of treatment-related complications associated with multiple myeloma. Journal of the American Medical Informatics Association : JAMIA. 2013;20(4):696-9.
458. Roesch SL, Styer AM, Wood GC, Kosak Z, Seiler J, Benotti P, et al. Perturbations of fibroblast growth factors 19 and 21 in type 2 diabetes. PloS one. 2015;10(2).
459. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics (Oxford, England). 2010;26(9):1205-10.
460. Cortes A, Dendrou CA, Motyer A, Jostins L, Vukcevic D, Dilthey A, et al. Bayesian analysis of genetic association across tree-structured routine healthcare data in the UK Biobank. Nature genetics. 2017;49(9):1311-8.
461. Li X, Meng X, Spiliopoulou A, Timofeeva M, Wei WQ, Gifford A, et al. MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank. Annals of the rheumatic diseases. 2018.
462. Liu R, Han C, Wu D, Xia X, Gu J, Guan H, et al. Prevalence of Hyperuricemia and Gout in Mainland China from 2000 to 2014: A Systematic Review and Meta-Analysis. Biomed Res Int. 2015;2015:762820.
463. Emma Smith LM. Global Prevalence of Hyperuricemia: A Systematic Review of Population-Based Epidemiological Studies. Arthritis Rheumatol. 2015;67(suppl 10).
464. Daniel I. Feig D-HK, Richard J. Johnson, . Uric Acid and Cardiovascular Risk. NEJM. 2009;359(17):1811-21.
465. Nejatinamini S, Ataie-Jafari A, Qorbani M, Nikoohemat S, Kelishadi R, Asayesh H, et al. Association between serum uric acid level and metabolic syndrome components. Journal of diabetes and metabolic disorders. 2015;14:70.
466. Li X, Meng X, Timofeeva M, Tzoulaki I, Tsilidis KK, Ioannidis PA, et al. Serum uric acid levels and multiple health outcomes: umbrella review of evidence from observational studies, randomised controlled trials, and Mendelian randomisation studies. BMJ (Clinical research ed). 2017;357:j2376.
467. Millard LA, Davies NM, Timpson NJ, Tilling K, Flach PA, Davey Smith G. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. Scientific reports. 2015;5:16645.
468. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. European journal of epidemiology. 2015;30(7):543-52.

469. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International journal of epidemiology. 2015;44(2):512-25.
470. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature genetics. 2016;48(5):481-7.
471. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. Bioinformatics (Oxford, England). 2014;30(16):2375-6.
472. Anurag Verma YB SD, Shefali S Verma, Sarah A Pendergrass, Marylyn D Ritchie. A simulation study investigating power estimates in Phenome-Wide Association Studies bioRxiv 115550;. 2017.
473. Benjamini Yoav HY. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. 1995;72(4):405-16.
474. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genetic epidemiology. 2013;37(7):658-65.
475. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nature genetics. 2009;41(6):666-76.
476. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nature genetics. 2015;47(10):1121-30.
477. Kullo IJ, Shameer K, Jouni H, Lesnick TG, Pathak J, Chute CG, et al. The ATXN2-SH2B3 locus is associated with peripheral arterial disease: an electronic medical record-based genome-wide association study. Frontiers in genetics. 2014;5:166.
478. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. Nature genetics. 2010;42(4):295-302.
479. Medici M, Porcu E, Pistis G, Teumer A, Brown SJ, Jensen RA, et al. Identification of novel genetic Loci associated with thyroid peroxidase antibodies and clinical thyroid disease. PLoS genetics. 2014;10(2):e1004123.
480. Cadzow M, Merriman TR, Dalbeth N. Performance of gout definitions for genetic epidemiological studies: analysis of UK Biobank. Arthritis research & therapy. 2017;19(1):181.
481. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. European journal of epidemiology. 2017.
482. Yoshitomi R, Fukui A, Nakayama M, Ura Y, Ikeda H, Oniki H, et al. Sex differences in the association between serum uric acid levels and cardiac hypertrophy in patients with chronic kidney disease. Hypertension research : official journal of the Japanese Society of Hypertension. 2014;37(3):246-52.
483. Li YH, Lin GM. Sex differences in the association between serum uric acid and cardiac mortality in Asian patients with established coronary artery disease: A revisit of the ET-CHD registry in Taiwan, 1997-2003. Journal of cardiology. 2016;68(5):461.
484. Rodrigues SL, Baldo MP, Capingana P, Magalhaes P, Dantas EM, Molina Mdel C, et al. Gender distribution of serum uric acid and cardiovascular risk factors: population based study. Arquivos brasileiros de cardiologia. 2012;98(1):13-21.

485. Kurata A, Shigematsu Y, Higaki J. Sex-related differences in relations of uric acid to left ventricular hypertrophy and remodeling in Japanese hypertensive patients. Hypertension research : official journal of the Japanese Society of Hypertension. 2005;28(2):133-9.
486. Feig DI KD-H, Johnson RJ. Uric Acid and Cardiovascular Risk The New England journal of medicine. 2008;359(17):1811-21.
487. Kawabe M, Sato A, Hoshi T, Sakai S, Hiraya D, Watabe H, et al. Gender differences in the association between serum uric acid and prognosis in patients with acute coronary syndrome. Journal of cardiology. 2016;67(2):170-6.
488. Kullo IJ, Ding K, Jouni H, Smith CY, Chute CG. A genome-wide association study of red blood cell traits using the electronic medical record. PloS one. 2010;5(9).
489. Pichler I, Minelli C, Sanna S, Tanaka T, Schwienbacher C, Naitza S, et al. Identification of a common variant in the TFR2 gene implicated in the physiological regulation of serum iron levels. Human molecular genetics. 2011;20(6):1232-40.
490. Pendergrass SA, Brown-Gentry K, Dudek S, Frase A, Torstenson ES, Goodloe R, et al. Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. PLoS genetics. 2013;9(1):e1003087.
491. Hall MA, Verma A, Brown-Gentry KD, Goodloe R, Boston J, Wilson S, et al. Detection of pleiotropy through a Phenome-wide association study (PheWAS) of epidemiologic data as part of the Environmental Architecture for Genes Linked to Environment (EAGLE) study. PLoS genetics. 2014;10(12):e1004678.
492. Devalliere J, Charreau B. The adaptor Lnk (SH2B3): an emerging regulator in vascular cells and a link between immune and inflammatory signaling. Biochemical pharmacology. 2011;82(10):1391-402.
493. Devalliere J, Chatelais M, Fitau J, Gerard N, Hulin P, Velazquez L, et al. LNK (SH2B3) is a key regulator of integrin signaling in endothelial cells and targets alpha-parvin to control cell adhesion and migration. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2012;26(6):2592-606.
494. Peden DB, Hohman R, Brown ME, Mason RT, Berkebile C, Fales HM, et al. Uric acid is a major antioxidant in human nasal airway secretions. Proceedings of the National Academy of Sciences of the United States of America. 1990;87(19):7638-42.
495. Kool M, Willart MA, van Nimwegen M, Bergen I, Pouliot P, Virchow JC, et al. An unexpected role for uric acid as an inducer of T helper 2 cell immunity to inhaled antigens and inflammatory mediator of allergic asthma. Immunity. 2011;34(4):527-40.
496. Hara K, Iijima K, Elias MK, Seno S, Tojima I, Kobayashi T, et al. Airway uric acid is a sensor of inhaled protease allergens and initiates type 2 immune responses in respiratory mucosa. Journal of immunology (Baltimore, Md : 1950). 2014;192(9):4032-42.
497. Webb R, Jeffries M, Sawalha AH. Uric acid directly promotes human T-cell activation. The American journal of the medical sciences. 2009;337(1):23-7.
498. Gold MJ, Hiebert PR, Park HY, Stefanowicz D, Le A, Starkey MR, et al. Mucosal production of uric acid by airway epithelial cells contributes to particulate matter-induced allergic sensitization. Mucosal immunology. 2016;9(3):809-20.
499. Horsfall LJ, Nazareth I, Petersen I. Serum uric acid and the risk of respiratory disease: a population-based cohort study. Thorax. 2014;69(11):1021-6.

500. Aida Y, Shibata Y, Osaka D, Abe S, Inoue S, Fukuzaki K, et al. The relationship between serum uric acid and spirometric values in participants in a health check: the Takahata study. International journal of medical sciences. 2011;8(6):470-8.
501. Cortes A, Dendrou C, Motyer A, Jostins L, Vukcevic D, Dilthey A, et al. Bayesian analysis of genetic association across tree-structured routine healthcare data in the UK Biobank. bioRxiv. 2017:105122.
502. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. bioRxiv. 2017:176834.
503. Latourte A, Soumare A, Bardin T, Perez-Ruiz F, Debette S, Richette P. Uric acid and incident dementia over 12 years of follow-up: a population-based cohort study. Annals of the rheumatic diseases. 2017.
504. Khan AA, Quinn TJ, Hewitt J, Fan Y, Dawson J. Serum uric acid level and association with cognitive impairment and dementia: systematic review and meta-analysis. Age (Dordrecht, Netherlands). 2016;38(1):16.
505. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336-7.
506. van der Harst P, Zhang W, Mateo Leach I, Rendon A, Verweij N, Sehmi J, et al. Seventy-five genetic loci influencing the human red blood cell. Nature. 2012;492(7429):369-75.
507. Draisma HH, Pool R, Kobl M, Jansen R, Petersen AK, Vaarhorst AA, et al. Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. Nature communications. 2015;6:7208.
508. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. Nature genetics. 2014;46(6):543-50.
509. Nakayama A, Nakaoka H, Yamamoto K, Sakiyama M, Shaukat A, Toyoda Y, et al. GWAS of clinically defined gout and subtypes identifies multiple susceptibility loci that include urate transporter genes. Annals of the rheumatic diseases. 2017;76(5):869-77.
510. Ligthart S, Vaez A, Hsu YH, Stolk R, Uitterlinden AG, Hofman A, et al. Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. BMC genomics. 2016;17:443.
511. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nature genetics. 2009;41(1):56-65.
512. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. Nature genetics. 2010;42(3):210-5.
513. Kottgen A, Pattaro C, Boger CA, Fuchsberger C, Olden M, Glazer NL, et al. New loci associated with kidney function and chronic kidney disease. Nature genetics. 2010;42(5):376-84.
514. Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, et al. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nature genetics. 2010;42(8):684-7.
515. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nature communications. 2016;7:10023.

516. Kettunen J, Demirkan A, Wurtz P, Draisma HH, Haller T, Rawal R, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nature communications. 2016;7:11122.
517. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nature genetics. 2008;40(4):395-402.
518. Zhernakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, Franke L, et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS genetics. 2011;7(2):e1002004.
519. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nature genetics. 2015;47(10):1121-30.
520. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nature genetics. 2015;47(9):979-86.
521. Zemunik T, Boban M, Lauc G, Jankovic S, Rotim K, Vatavuk Z, et al. Genome-wide association study of biochemical traits in Korcula Island, Croatia. Croat Med J. 2009;50(1):23-33.
522. Mazzali M, Kanbay M, Segal MS, Shafiu M, Jalal D, Feig DI, et al. Uric acid and hypertension: cause or effect? Current rheumatology reports. 2010;12(2):108-17.
523. Vitart V, Rudan I, Hayward C, Gray NK, Floyd J, Palmer CN, et al. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. Nature genetics. 2008;40(4):437-42.
524. Gibran Hemani, Jie Zheng, Kaitlin H Wade, Charles Laurin, Benjamin Elsworth, Stephen Burgess, et al. MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. bioRxiv doi: https://doiorg/101101/078972. 2016.
525. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518(7538):187-96.
526. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nature genetics. 2013;45(11):1274-83.
527. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nature genetics. 2012;44(6):659-69.
528. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011;478(7367):103-9.
529. Martinon F, Petrilli V, Mayor A, Tardivel A, Tschopp J. Gout-associated uric acid crystals activate the NALP3 inflammasome. Nature. 2006;440(7081):237-41.
530. Martinon F. Mechanisms of uric acid crystal-mediated autoinflammation. Immunological reviews. 2010;233(1):218-32.
531. Ding B, Padyukov L, Lundstrom E, Seielstad M, Plenge RM, Oksenberg JR, et al. Different patterns of associations with anti-citrullinated protein antibody-positive and anti-

citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region. Arthritis and rheumatism. 2009;60(1):30-8.
532. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. The New England journal of medicine. 2011;365(23):2205-19.
533. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014;506(7488):376-81.
534. Wang J, Qin T, Chen J, Li Y, Wang L, Huang H, et al. Hyperuricemia and risk of incident hypertension: a systematic review and meta-analysis of observational studies. PloS one. 2014;9(12):e114259.
535. Billiet L, Doaty S, Katz JD, Velasquez MT. Review of hyperuricemia as new marker for metabolic syndrome. ISRN rheumatology. 2014;2014:852954.
536. Peng TC, Wang CC, Kao TW, Chan JY, Yang YH, Chang YW, et al. Relationship between hyperuricemia and lipid profiles in US adults. Biomed Res Int. 2015;2015:127596.
537. Li C, Hsieh MC, Chang SJ. Metabolic syndrome, diabetes, and hyperuricemia. Current opinion in rheumatology. 2013;25(2):210-6.
538. Raimondo A, Rees MG, Gloyn AL. Glucokinase regulatory protein: complexity at the crossroads of triglyceride and glucose metabolism. Current opinion in lipidology. 2015;26(2):88-95.
539. Neogi T. Gout and Calcium Crystal Related Arthropathies: An Issue of Rheumatic Disease Clinics. Web: E-Book; 2014.
540. Mazharian A, Mori J, Wang YJ, Heising S, Neel BG, Watson SP, et al. Megakaryocyte-specific deletion of the protein-tyrosine phosphatases Shp1 and Shp2 causes abnormal megakaryocyte development, platelet production, and function. Blood. 2013;121(20):4205-20.
541. Combe B, Landewe R, Daien CI, Hua C, Aletaha D, Alvaro-Gracia JM, et al. 2016 update of the EULAR recommendations for the management of early arthritis. Annals of the rheumatic diseases. 2017;76(6):948-59.
542. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. Nature reviews Genetics. 2016;17(3):129-45.
543. Robinson PN, Mungall CJ, Haendel M. Capturing phenotypes for precision medicine. Cold Spring Harbor molecular case studies. 2015;1(1):a000372.
544. World Health Organization. International Classification of Diseases (ICD) 2016 [cited 2018 21 May]. Available from: http://www.who.int/classifications/icd/en/.
545. World Health Organization. ICD-10 Version:2015 [cited 2018 21 May]. Available from: http://apps.who.int/classifications/icd10/browse/2015/en.
546. Min KB, Min JY. Increased risk for hyperuricemia in adults sensitized to cow milk allergen. Clinical Rheumatology. 2017;36(6):1407-12.
547. Centers for Medicare & Medicaid Services. 2015 ICD-10-CM and GEMs. [cited 2018 21 May]. Available from: https://www.cms.gov/medicare/coding/icd10/2015-icd-10-cm-and-gems.html.
548. Hebbring SJ. The challenges, advantages and future of phenome-wide association studies. Immunology. 2014;141(2):157-65.
549. Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, et al. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. Nucleic acids research. 2015;43(Database issue):D1071-8.

550. Kohler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic acids research. 2014;42(Database issue):D966-74.
551. Sioutos N, de Coronado S, Haber MW, Hartel FW, Shaiu WL, Wright LW. NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. Journal of biomedical informatics. 2007;40(1):30-43.
552. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Studies in health technology and informatics. 2006;121:279-90.
553. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic acids research. 2004;32(Database issue):D267-70.
554. Amberger J, Bocchini C, Hamosh A. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM(R)). Human mutation. 2011;32(5):564-7.
555. Pendergrass SA, Ritchie MD. Phenome-Wide Association Studies: Leveraging Comprehensive Phenotypic and Genotypic Data for Discovery. Current genetic medicine reports. 2015;3(2):92-100.
556. Li R, Huang C, Chen J, Guo Y, Tan S. The role of uric acid as a potential neuroprotectant in acute ischemic stroke: a review of literature. Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology. 2015;36(7):1097-103.
557. Chiquete E, Ruiz-Sandoval JL, Murillo-Bonilla LM, Arauz A, Orozco-Valera DR, Ochoa-Guzman A, et al. Serum uric acid and outcome after acute ischemic stroke: PREMIER study. Cerebrovasc Dis. 2013;35(2):168-74.
558. Gerber Y, Tanne D, Medalie JH, Goldbourt U. Serum uric acid and long-term mortality from stroke, coronary heart disease and all causes. European journal of cardiovascular prevention and rehabilitation : official journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology. 2006;13(2):193-8.
559. Smith GD ES. Mendelian Randomization: Genetic Variants as Instruments for Strengthening Causal Inference in Observational Studies. . Washington (DC): National Academies Press (US) Available from: https://wwwncbinlmnihgov/books/NBK62433/. 2006;16.
560. Sanderson E, Windmeijer F. A weak instrument [Formula: see text]-test in linear IV models with multiple endogenous variables. Journal of econometrics. 2016;190(2):212-21.
561. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Statistical methods in medical research. 2017;26(5):2333-55.
562. Palmer TM, Holmes MV, Keating BJ, Sheehan NA. Correcting the Standard Errors of 2-Stage Residual Inclusion Estimators for Mendelian Randomization Studies. American journal of epidemiology. 2017;186(9):1104-14.
563. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Statistics in medicine. 2008;27(8):1133-63.
564. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Statistics in medicine. 1999;18(20):2693-708.
565. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to conditioning on a collider. International journal of epidemiology. 2010;39(2):417-20.

566. Robinson PC, Merriman TR, Herbison P, Highton J. Hospital admissions associated with gout and their comorbidities in New Zealand and England 1999-2009. Rheumatology (Oxford, England). 2013;52(1):118-26.
567. Sharon Y, Schlesinger N. Beyond Joints: a Review of Ocular Abnormalities in Gout and Hyperuricemia. Current rheumatology reports. 2016;18(6):37.
568. McCarty CA, Nanjan MB, Taylor HR. Attributable risk estimates for cataract to prioritize medical and public health action. Investigative ophthalmology & visual science. 2000;41(12):3720-5.